NOVEL SECRETED PROTEINS

FIELD OF THE INVENTION

The invention relates to novel secreted polypeptide species encoded by genomic sequences previously thought to be noncoding regions, which are differentially expressed in individuals with cardiovascular disorders. The invention also relates to isolated polynucleotides encoding such polypeptides, polymorphic variants thereof, and the use of said nucleic acids and polypeptides or compositions thereof in detection assays, for cardiovascular disorder diagnosis.

BACKGROUND

In the past few years, the sequencing of the human genome and extensive datamining of expressed sequence tags (ESTs), combined with powerful bioinformatic tools for analysis and prediction have provided the field of biotechnology with a wealth of information about the structure of the genome. For example, certain markers such as promoter sequences, splice sites, and polyA tail sequences are used to detect the presence of coding regions in the genome. Several exon identification methods, some with gene assembly capabilities, have been developed. These include Markov and Hidden Markov models, e.g. P. Baldi, et al., Proc. Nat. Acad. Sci., 91:1059-1063; statistical methods, e.g. R. Guigo, et al., J.Mol.Biol. 226:141-157 (1992); homology, e.g. W. Pearson, et al., Proc. Nat. Acad. Sci., 85:2444-2448; fourier transform analysis, e.g. Yan, et al. (1998), Bioinformatics, 14:685-690; as well as neural networks, e.g. E. Uberbacher, et al., Proc. Nat. Acad. Sci., 88: 11261-11265; and game theory, e.g. Jeffrey, H. Nucleic Acids Res. 13:3453-3462. Additional methods for detecting coding sequences are disclosed in US Patent 60094626 from Vanderbilt Universiy and WO 01/16861 from Genetics Institute. From this information, it is theoretically possible to predict the sequence of the protein(s) encoded by every coding region in the genome. However, the current prediction systems do have inherent weaknesses, such as reliance on statistical data collected from previously characterized sequences.

The present invention provides a method of detecting the full range of the proteome of secreted proteins. It relies on a system and methods for identifying biomolecules actually present in a biological sample, for example, protein markers. More specifically, the present invention relies on protein fractionation of samples where large volumes of biological fluid samples are analyzed to identify proteins present in a wide range of concentrations. The analysis of a proteome involves the separation of the proteins in a sample followed by the identification of the resolved proteins. For complex samples such as human plasma, this is a challenging task in view of the large numbers of proteins expected to be present and the wide dynamic range of concentrations, known to span at least

11 to 12 orders of magnitude. In order to identify and characterize polypeptides present at very low concentrations, it is necessary to start with a large sample volume to ensure a sufficient quantity for detection by mass spectrometry.

Using the methods of the invention, the inventors have identified proteins encoded by genomic sequences that have previously been described as noncoding sequences. Thus, the invention provides not only new proteins with diagnostic application, but provides new insight into the mechanism of protein expression and the potential of the human genome.

Cardiovascular disease is a major health risk throughout the industrialized world. Coronary Artery Disease (CAD) is characterized by atherosclerosis or hardening of the arteries. Atherosclerosis is the most prevalent of cardiovascular diseases, is the principal cause of heart attack, stroke, and gangrene of the extremities, and thereby the principle cause of death in the United States. Atherosclerosis is a complex disease involving many cell types and molecular factors (described in, for example, Ross, 1993, Nature 362: 801-809). In normal circumstances a protective response to insults to the endothelium and smooth muscle cells (SMCs) of the wall of the artery consists of the formation of fibrofatty and fibrous lesions or plaques, preceded and accompanied by inflammation. The advanced lesions of atherosclerosis may occlude the artery concerned, and result from an excessive inflammatory-fibroproliferative response to numerous different forms of insult. Injury or dysfunction of the vascular endothelium is a common feature of many conditions that predispose an individual to accelerated development of atherosclerotic cardiovascular disease.

Atherosclerotic plaques occlude the blood vessel concerned and restrict the flow of blood, resulting in ischemia. Ischemia is a condition characterized by a lack of oxygen supply in tissues of organs due to inadequate perfusion. Such inadequate perfusion can have a number of natural causes, including atherosclerotic or restenotic lesions, anemia, or stroke. The most common cause of ischemia in the heart is atherosclerotic disease of epicardial coronary arteries. By reducing the lumen of these vessels, atherosclerosis causes an absolute decrease in myocardial perfusion in the basal state or limits appropriate increases in perfusion when the demand for flow is augmented. Coronary blood flow can also be limited by arterial thrombi, spasm, and, rarely, coronary emboli, as well as by ostial narrowing due to luetic aortitis. Congenital abnormalities, such as anomalous origin of the left anterior descending coronary artery from the pulmonary artery, may cause myocardial ischemia and infarction in infancy, but this cause is very rare in adults.

Myocardial ischemia can also occur if myocardial oxygen demands are abnormally increased, as in severe ventricular hypertrophy due to hypertension or aortic stenosis. The latter can be present with angina that is indistinguishable from that caused by coronary atherosclerosis. A reduction in the oxygen-carrying capacity of the blood, as in extremely severe anemia or in the presence of carboxy-hemoglobin, is a rare cause of myocardial ischemia. Not infrequently, two or more causes of ischemia

will coexist, such as an increase in oxygen demand due to left ventricular hypertrophy and a reduction in oxygen supply secondary to coronary atherosclerosis.

Extensive clinical studies have identified factors that increase the risk of cardiovascular disorders. Some of these risk factors, such as age, gender, and family history cannot be changed. Other risk factors include the following: smoking, high blood pressure, high fat and high cholesterol diet, diabetes, lack of exercise, obesity, and stress.

Fortunately, many contributing factors are controllable through lifestyle changes. The risk of cardiovascular disorders for smokers is more than twice that of non-smokers. When a person stops smoking, regardless of how much he or she may have smoked in the past, their risk of developing a disorder rapidly declines. Serum cholesterol level is directly related to prevalence of cardiovascular disorder and hypertension or high blood pressure is an important risk factor. Physical activity has been postulated to reduce the risk of developing a cardiovascular disorder through various mechanisms: it increases myocardial oxygen supply, decreases oxygen demand, and improves myocardial contraction and its electrical impulse stability. Reduced oxygen demand and myocardial work are reflected in lowered heart rate and blood pressure at rest. Physical activity also increases the diameter and dilatory capacity of coronary arteries, increases collateral artery formation, and reduces rates of progression of coronary artery atherosclerosis. Obesity and the serum fatty acids are reduced by activity.

There may be no noticeable symptoms of a cardiovascular disorder at rest, but symptoms such as chest pressure may occur with increased activity or stress. Other first signs that can appear are heartburn, nausea, vomiting, numbness, shortness of breath, heavy cold sweating, unexplained fatigue, and feelings of anxiety. The more severe symptoms of cardiovascular disorders are chest pain (angina pectoris), rhythm disturbances (arrhythmias), stroke, or heart attack (myocardial infarction). Strokes and heart attacks result from a blocked artery in the brain and heart tissue, respectively. Because symptoms vary, the tests and treatments chosen can be very different from one patient to another.

Diagnostic tests useful in determining the extent and severity of cardiovascular disorder include: electrocardiogram (EKG), stress test, nuclear scanning, coronary angiography, resting EKG, EKG Multiphase Information Diagnosis Indexes, Holter monitor, late potentials, EKG mapping, echocardiogram, Thallium scan, PET, MRI, CT, angiogram and IVUS. Additional risk factor measures and useful diagnostics are common and best applied by one of skill in the art of medicine. There are many different therapeutic approaches, depending on the seriousness of the disease. For many people, cardiovascular disorders are managed with lifestyle changes and medications. More severe diagnoses may indicate a need for surgery.

Surgical approaches to the treatment of ischemic atherosclerosis include bypass grafting, coronary angioplasty, laser angioplasty, atherectomy, endarterectomy, and percutaneous translumenal

angioplasty (PCTA). The failure rate after these approaches due to restenosis, in which the occlusions recur and often become even worse, is extraordinarily high (30-50%). It appears that much of the restenosis due to further inflammation, smooth muscle accumulation, and thrombosis. Additional therapeutic approaches to cardiovascular disease have included treatments that encouraged angiogenesis in such conditions as ischemic heart and limb disease.

The non-specific nature of most CAD and cardiovascular disorder symptoms makes definitive diagnosis difficult. More quantitative diagnostic methods suffer from variability, both between individuals and between readings on a single individual. Thus, diagnostic measures must be standardized and applied to individuals with well-documented and extensive medical histories. Further, current diagnostic methods often do not reveal the underlying cause for a given observation or reading. Therefore, a therapeutic strategy based on a particular positive result likely will not address the causative problem and may even be harmful to the individual.

Methods of diagnosis that rely on nucleotide detection include genetic approaches and expression profiling. For example, genes that are known to be involved in cardiovascular disorders may be screened for mutations using common genotyping techniques such as sequencing, hybridization-based techniques, or PCR. In another example, expression from a known gene may be tracked by standard techniques including RTPCR, various hybridization-based techniques, and sequencing. These strategies often do not enable a practitioner to detect differences in mRNA processing and splicing, translation rate, mRNA stability, and posttranslational modifications such as proteolytic processing, phosphorylation, glycosylation, and amidation.

To address the current weaknesses in the diagnostic state of the art for cardiovascular disorders, the invention provides specific polypeptides that are differentially expressed in plasma from individuals with Coronary Artery Disease compared to control plasma. By providing the actual polypeptide species, differences in mRNA processing and splicing, translation rate, mRNA stability, and posttranslational modifications such as proteolytic processing, phosphorylation, glycosylation, and amidation are revealed.

To this end, the polypeptides of the invention are described as "Novel Plasma Polypeptides" or NPPs. These polypeptide sequences are described as comprising at least one of the amino acid sequences selected from the peptides of Table 1 (SEQ ID NOs:1-106). Full length polypeptides corresponding to selected peptides are described as SEQ ID NOs:107-122. NPP-encoding polynucleotides represent novel coding sequences and are presented as SEQ ID NOs:123-138.

The present invention discloses "Novel Plasma Polypeptides" (NPPs), fragments, and post-translationally modified species of NPPs that are present at a different (i.e., increased or decreased) level in plasma obtained from individuals with Coronary Artery Disease (CAD). Thus, the NPPs of the invention represent an important diagnostic tool for determining the risk of CAD, coronary heart

disease (CHD), peripheral vascular disease, cerebral ischemia (stroke), congestive heart failure, atherosclerosis, hypertension, and other cardiovascular diseases. NPPs are secreted factors and as such, are easy to target, e.g., with a detectable molecule or antibody. Thus, the polypeptide species of the invention are useful for diagnosis of cardiovascular disease.

SUMMARY OF THE INVENTION

The present invention is directed to compositions related to secreted polypeptides, designated herein "Novel Plasma Polypeptides" or "NPPs". Such compositions include the NPPs, having an amino acid sequence of SEQ ID NO:1-106, NPP proteins of SEQ ID NOs:107-122, NPP precursors, NPP-encoding polynucleotides, NPP antibodies, including monoclonal antibodies and other binding compositions derived therefrom, and methods of making and using these compositions. NPP precursors of the invention include the NPP polypeptides of SEQ ID NOs:107-122. NPP-encoding polynucletides are described as SEQ ID NOs:123-138.

A preferred embodiment of the invention includes NPPs having a posttranslational modification, such as a phosphorylation, glycosylation, acetylation, amidation, or a C-, N- or O-linked carbohydrate group. Additionally preferred are NPPs with intra- or inter-molecular interactions, e.g., disulfide and hydrogen bonds, that result in higher order structures. Also preferred are NPPs that result from differential mRNA processing or splicing.

In another aspect, the invention includes isolated polynucleotides coding for a polypeptide comprising an amino acid sequence of one of SEQ ID NOs:1-122, antisense oligonucleotides complementary to such sequences, oligonucleotides complementary to NPP gene sequences useful in diagnostic and analytical assays, such as primers for polymerase chain reactions (PCRs), and vectors for expressing NPP peptides. In particular, isolated polynucleotides corresponding to the sequences of SEQ ID NOs:123-138, sequences complementary to SEQ ID NOs:123-138, and fragments thereof (described herein) are included in the invention.

In another aspect, the invention includes NPPs having a sequence which is at least 80 or 85, preferably at least 90 or at least 95 percent identical to a sequence selected from SEQ ID NOs:1-122. Most preferably, the invention includes polypeptides having at least 97 percent, and more preferably at least 98 percent, and still more preferably at least 99 percent, identity with a sequence selected from SEQ ID NO:1-122.

In an additional aspect, the invention includes modified NPPs. Such modifications include protecting/blocking groups, linkage to an antibody molecule or other cellular ligand, and detectable labels, such as an enzymatic, fluorescent, isotopic or affinity label to allow for detection and isolation of the protein. Chemical modifications may be carried out by known techniques, including but not

limited, to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH4, acetylation, formylation, oxidation, reduction, or metabolic synthesis in the presence of tunicamycin. Also provided by the invention are chemically modified derivatives of the polypeptides of the invention which may provide additional advantages such as increased solubility, stability and circulating time of the polypeptide, or decreased immunogenicity (e.g., water soluble polymers such as polyethylene glycol, ethylene glycol/propylene glycol copolymers, carboxymethylcellulose, dextran, polyvinyl alcohol). NPPs are modified at random positions within the molecule, or at predetermined positions within the molecule and may include one, two, three or more attached chemical moieties.

A preferred aspect of the invention provides a composition comprising an isolated NPP, i.e., a NPP free from proteins or protein isoforms having a significantly different isoelectric point or a significantly different apparent molecular weight from the NPP. The isoelectric point and molecular weight of a NPP may be indicated by affinity and size-based separation chromatography, 2-dimensional gel analysis, and mass spectrometry. Preferrably, the composition comprises two or more isolated NPPs.

In a preferred aspect, the invention provides particular polypeptide species that comprise an amino acid sequence selected from the group consisting of SEQ ID NOs:18, 20, 27, 30, 31, 43, 47, 53, 55, 62, 66, 67, 73, 76, 96, and 102. Preferably, the particular polypeptide species further comprises contiguous amino acid sequence from the corresponding full length polypeptide selected from SEQ ID NOs:107-122 (see Table 1). Preferred species are polypeptides that i) comprise an amino acid sequence selected from the group consisting of SEQ ID NOs: 18, 20, 27, 30, 31, 43, 47, 53, 55, 62, 66, 67, 73, 76, 96, and 102; ii) appear in human blood plasma; and iii) result from proteolytic processing of the corresponding full length polypeptide of SEQ ID NO:107-122. Especially preferred are the peptides of SEQ ID NOs:18, 20, 27, 31, 47, 53, 62, 66, 73, and 76, corresponding to the preferred full length proteins of SEQ ID NOs:107-109, 111, 113, 114, 116, 117, 119, and 120.

In another aspect, the invention includes isolated antibodies capable of binding any of the polypeptides, peptide fragments, or peptides described above. Preferably, the antibodies of the invention are monoclonal antibodies. Further preferred are antibodies that bind to a NPP specifically, that is, antibodies that do not recognize other polypeptides with high affinity. Anti-NPP antibodies have purification and diagnostic applications, particularly for NPP-related disorders. Preferred anti-NPP protein antibodies for purification and diagnosis are attached to a label group. Preferred NPP-related disorders for diagnosis include coronary artery disease (CAD), coronary heart disease (CHD), peripheral vascular disease, cerebral ischemia (stroke), congestive heart failure, atherosclerosis, hypertension, and other cardiovascular diseases. Diagnostic methods include, but are not limited to, those that employ antibodies or antibody-derived compositions specific for a NPP antigen.

Diagnostic methods for detecting NPPs in specific tissue samples and biological fluids (preferably plasma), and for detecting levels of expression of NPPs in tissues, also form part of the invention. Compositions comprising one or more antibodies described above, together with a pharmaceutically acceptable carrier are also within the scope of the invention, for example, for in vivo diagnosis methods.

The invention further includes methods of using NPP-related compositions, including primers complementary to the NPP genes and/or messenger RNA and anti-NPP antibodies, for detecting and measuring quantities of the NPPs in tissues and biological fluids, preferably blood plasma. The invention provides methods for diagnosis of cardiovascular disorders that comprise detecting the level of at least one NPP in a sample of body fluid, preferably blood plasma. These methods are also suitable for clinical screening, prognosis, monitoring the results of therapy, identifying patients most likely to respond to a particular therapeutic treatment, and identifying new targets for drug treatment.

In another embodiment, the invention provides a method of identifying a modulator of at least one NPP biological activity comprising the steps of: i) contacting a test modulator of a NPP biological activity with the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NOs:1-122; ii) detecting the level of said NPP biological activity; and iii) comparing the level of said NPP biological activity to that of a control sample lacking said test modulator. Where the difference in the level of NPP protein biological activity is a decrease, the test modulator is an inhibitor of at least one NPP biological activity. Where the difference in the level of NPP biological activity is an increase, the test substance is an activator of at least one NPP biological activity. Preferably, the method comprises the detection of the level or the assessing of at least one biological activity of two or more NPP proteins.

Another aspect of the invention provides a method of identifying a modulator of a cardiovascular disorder comprising the steps of (a) administering a candidate agent to a non-human test animal which is predisposed to be affected or which is affected by the cardiovascular disorder; (b) administering the candidate agent of (a) to a matched control non-human animal not predisposed to be affected or not being affected by the cardiovascular disorder; (c) detecting and /or quantifying the level of at least one polypeptide in a biological sample obtained from the non-human test animal of step (a) and from the control animal of step (b), wherein the at least one polypeptide is selected from (i) a polypeptide selected from the group consisting of SEQ ID NOs:1-122; (ii) a polypeptide comprising an amino acid sequence which shares more than about 60% but less than 100% homology with the amino acid sequence selected from the group consisting of SEQ ID NOs:1-122; and (iii) a polypeptide comprising at least 6 contiguous amino acids of a protein sequence as defined in i) or ii); and step (d) comparing the levels of the at least one polypeptide of step (c); wherein a displacement of the level of the at least one polypeptide in the biological sample obtained from the non-human test

animal towards the level of the at least one polypeptide in the biological sample obtained from the control animal indicates that the candidate agent is a modulator of the cardiovascular disorder. In a preferred embodiment of the invention, the non-human test animal which is predisposed to be affected or which is affected by the cardiovascular disorder comprises an altered plasma level of at least one polypeptide selected from (i) a polypeptide selected from the group consisting of SEQ ID NOs:1-122; (ii) a polypeptide comprising an amino acid sequence which shares more than about 60% but less than 100% homology with the amino acid sequence selected from the group consisting of SEQ ID NOs:1-122; and (iii) a polypeptide comprising at least 6 contiguous amino acids of a protein sequence as defined in i) or ii).

Another aspect of the invention related to a method for monitoring the efficacy of a treatment of a subject having or at risk of developing a cardiovascular disorder with an agent, the method comprising the steps (a) obtaining a pre-administration biological sample from the subject prior to administration of the agent; (b) detecting and /or quantifying the level of at least one polypeptide in the biological sample from said subject, wherein the at least one polypeptide is selected from (i) a polypeptide selected from the group consisting of SEQ ID NOs:1-122; (ii) a polypeptide comprising an amino acid sequence which shares more than about 60% but less than 100% homology with the amino acid sequence selected from the group consisting of SEQ ID NOs:1-122; and (iii) a polypeptide comprising at least 6 contiguous amino acids of a protein sequence as defined in i) or ii); and steps (c) obtaining one or more post-administration biological samples from the subject; (d) detecting the level of the at least one polypeptide in the post-administration sample or samples; (e) comparing the level of the at least one polypeptide in the pre-administration sample with the level of the polypeptide in the post- administration sample; and step (f) adjusting the administration of the agent accordingly.

The invention provides kits that may be used in the above-recited methods and that may comprise single or multiple preparations, or antibodies, together with other reagents, label groups, substrates, if needed, and directions for use. The kits may be used for diagnosis of disease.

In one embodiment, Coronary Artery Disease (CAD) is defined by the appearance of at least one symptom. Such symptoms become more serious as the disease progresses. CAD is often accompanied by reduced left ventricle capacity or output. Early CAD symptoms include elevated plasma levels of cholesterol and low-density lipoprotein (especially oxidized forms), as well as platelet-rich plasma aggregations. The vascular endothelium responds to inflammation and thus formation of plaques and levels of inflammatory and fibrinogenic factors increase. In addition, CAD, or atherosclerosis, is characterized by vascular calcification and hardening of the arteries. The resulting partial occlusion of the blood vessels leads to hypertension and ischemic heart disease. Eventual complete vascular occlusion results in myocardial infarction, stroke, or gangrene.

In a preferred embodiment, detection of a difference in plasma levels of at least one NPP of the invention between a tested and control individual indicates an increased risk that the tested individual will develop CAD. Preferably, said detection indicates that an individual has at least a 1.05-fold, 1.1-fold, 1.15-fold, and more preferably at least a 1.2-fold increased likelihood of developing CAD. Alternatively, detection of a difference in plasma levels of at least one NPP of the invention indicates that the tested individual has CAD. Preferably, more than 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, or 20 different NPPs are detected. The amount of NPP difference observed in a tested individual compared to a control sample will correlate with the certainty of the prediction or diagnosis of CAD. As individual plasma NPP levels will vary depending on family history and other risk factors, each will preferably be examined on a case-by-case basis. In preferred embodiments, NPP is detected in a human plasma sample by the methods of the invention. Especially preferred techniques are mass spectrometry and immunodetection. Preferably, a prediction or diagnosis of CAD is based on at least a 1.1-, 1.15-, 1.2-, 1.25-, and more preferably a 1.5-fold difference in the tested NPP level as compared to the control. Most preferably, the detection and /or the quantifying of the level of the NPP in a biological sample is performed ex vivo.

In another aspect, the invention includes primer pairs for carrying out a PCR to amplify a segment of a polynucleotide of the invention. Each primer of a pair is an oligonucleotide having a length of between 15 and 30 nucleotides such that i) one primer of the pair forms a perfectly matched duplex with one strand of a polynucleotide of the invention and the other primer of the pair forms a perfectly matched duplex with the complementary strand of the same polynucleotide, and ii) the primers of a pair form such perfectly matched duplexes at sites on the polynucleotide that are separated by a distance of between 10 and 2500 nucleotides. Preferably, the annealing temperature of each primer of a pair with its respective complementary sequence is substantially the same.

In another aspect, the invention includes natural variants of the NPPs having a frequency in a selected population of at least two percent. More preferably, such natural variant has a frequency in a selected population of at least five percent, and still more preferably, at least ten percent. Most preferably, such natural variant has a frequency in a selected population of at least twenty percent. The selected population may be any recognized population of study in the field of population genetics. Preferably, the selected population is Caucasian, Negroid, or Asian. More preferably, the selected population is French, German, English, Spanish, Swiss, Japanese, Chinese, Irish, Korean, Singaporean, Icelandic, North American, Israeli, Arab, Turkish, Greek, Italian, Polish, Pacific Islander, Finnish, Norwegian, Swedish, Estonian, Austrian, or Indian. More preferably, the selected population is Icelandic, Saami, Finnish, French of Caucasian ancestry, Swiss, Singaporean of Chinese ancestry, Korean, Japanese, Quebecian, North American Pima Indians, Pennsylvanian Amish and Amish Mennonite, Newfoundlander, or Polynesian.

In another aspect, the invention provides a vector comprising DNA encoding a NPP. The invention also includes host cells and transgenic nonhuman animals comprising such a vector. There is also provided a method of making a NPP or NPP precursor. One preferred method comprises the steps of (a) providing a host cell containing an expression vector as disclosed above; (b) culturing the host cell under conditions whereby the DNA segment is expressed; and (c) recovering the protein encoded by the DNA segment. Another preferred method comprises the steps of: (a) providing a host cell capable of expressing a NPP; (b) culturing said host cell under conditions that allow expression of said NPP; and (c) recovering said NPP. Within one embodiment the expression vector further comprises a secretory signal sequence operably linked to the DNA segment, the cell secretes the protein into a culture medium, and the protein is recovered from the medium. An especially preferred method of making a NPP includes chemical synthesis using standard peptide synthesis techniques, as described in the section titled "Chemical Manufacture of NPP compositions".

Further aspects of the invention are also described in the specification and in the claims.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

SEQ ID NOs:1-106 describe the amino acid sequences of peptides discovered according to the methods of the invention in human plasma.

SEQ ID NOs:107-122 describe novel full length amino acid sequences that correspond to the peptide sequences of SEQ ID NOs: 18, 20, 27, 30, 31, 43, 47, 53, 55, 62, 66, 67, 73, 76, 96, and 102, respectively. The first column of Table 1 provides the matching peptide and full length polypeptide SEQ ID NOs.

SEQ ID NOs:123-138 describe the sequences of the novel coding regions discovered according to the methods of the invention. The polynucleotides of SEQ ID NOs:123-138 encode the polypeptides of SEQ ID NOs:107-122, respectively.

BRIEF DESCRIPTION OF THE TABLES

Table 1 lists the trypic peptides present at different levels in plasma of individuals with Coronary Artery Disease (CAD) compared to controls. The SEQ ID NO of the tryptic peptide is given, with the SEQ ID NO of the corresponding protein and cDNA, if applicable, in parentheses. The NCBI accession number (19 July 2001 version) and the translation frame of the NPP are indicated. For peptides translated in frames 1-3, the start and end nucleotide position of the coding sequence are given relative to the start position of the corresponding NCBI polynucleotide sequence. The start and end nucleotide positions for the peptides translated in frames 4-6 are given relative to the end position

of the corresponding NCBI polynucleotide sequence. The sample in which each peptide was found is indicated in the Proteome column (Control or Coronary Artery Disease plasma). Olav scores are shown in the far right column.

Table 2 describes the purification conditions for the NPPs separated according to the protocol of Example 2. The column labelled CEX indicates in which of the 18 cation exchange fractions the tryptic peptide was eluted, and the column labelled Salt indicates the NaCl concentration (mM) for the elution of these fractions, according to the protocol described in Step 3 of Example 2 herein. RP1 refers to the reverse phase fraction (fractions 1-30), and %B indicates the percentage of elution buffer for these fractions, according to the protocol described in Step 4 of Example 2 herein. The reverse phase fraction (fractions 1-24) is indicated as the last two digits of the Run Number.

Table 3 describes the purification conditions for the NPPs separated according to the protocol of Example 3. The columns labelled Benzamidine- Red Sepharose, SCX-SAX, and Rotofor indicate the fractions in which the NPP was found.

An explanation of how to interpret these tables is provided in the section titled "Characterization of NPPs".

BRIEF DESCRIPTION OF THE FIGURES

Figures 1 to 16 show the results of a gene prediction analysis for each of the 16 selected peptides, indicating the translation frame, predicted exon positions relative to the polynucleotide sequence of the given NCBI accession number, and the matching full length polypeptide sequence. HMMgene predictions display polypeptides encoded by open reading frames within the indicated NCBI sequence entry. The tryptic peptide sequence is listed next to the polynucleotide positions of the predicted exon in which it is found. The full length polypeptide encoded by each ORF is shown and the peptide sequence is highlighted in bold and underlined. Figures 1 to 16 correspond to the peptide sequences of SEQ ID NOs:18, 20, 27, 31, 47, 53, 62, 66, 73, 76, 30, 43, 55, 96, 102, and 67 respectively.

DETAILED DESCRIPTION OF THE INVENTION

The present invention described in detail below provides novel peptide sequences discovered in human plasma, corresponding polynucleotide and full length polypeptide sequences, antibodies, and related methods. The invention also provides methods, compositions, and kits useful for screening and diagnosis of a cardiovascular disorder in a mammalian individual; for identifying individuals most likely to respond to a particular therapeutic treatment; for monitoring the results of cardiovascular disorder therapy; and for screening NPP modulators. For clarity of disclosure, and not

by way of limitation, the invention will be described with respect to the analysis of blood plasma samples. However, as one skilled in the art will appreciate, the assays and techniques described below can be applied to other biological fluid samples (e.g. cerebrospinal fluid, lymph, bile, serum, saliva or urine) or tissue samples from an individual at risk of having or developing a cardiovascular disorder. The methods and compositions of the present invention are useful for screening, diagnosis and prognosis of a living individual, but may also be used for postmortem diagnosis in an individual, for example, to identify family members who are at risk of developing the same disorder.

Definitions

As used herein, the term "nucleic acids" and "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. Throughout the present specification, the expression "nucleotide sequence" may be employed to designate indifferently a polynucleotide or a nucleic acid. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule. Also, used interchangeably herein are terms "nucleic acids", "oligonucleotides", and "polynucleotides".

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NPP nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. A nucleic acid molecule of the present invention can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid, as a hybridization probe, NPP nucleic acid molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular Cloning. A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively,

genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NPP nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

As used herein, the term "hybridizes to" is intended to describe conditions for moderate stringency or high stringency hybridization, preferably where the hybridization and washing conditions permit nucleotide sequences at least 60% homologous to each other to remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85%, 90%, 95% or 98% homologous to each other typically remain hybridized to each other. Stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. In a preferred, non-limiting example, stringent hybridization conditions are as follows: the hybridization step is realized at 65°C in the presence of 6 x SSC buffer, 5 x Denhardt's solution, 0,5% SDS and 100µg/ml of salmon sperm DNA. The hybridization step is followed by four washing steps:

- two washings during 5 min, preferably at 65°C in a 2 x SSC and 0.1%SDS buffer;
- one washing during 30 min, preferably at 65°C in a 2 x SSC and 0.1% SDS buffer,
- one washing during 10 min, preferably at 65°C in a 0.1 x SSC and 0.1%SDS buffer,

these hybridization conditions being suitable for a nucleic acid molecule of about 20 nucleotides in length. It will be appreciated that the hybridization conditions described above are to be adapted according to the length of the desired nucleic acid, following techniques well known to the one skilled in the art, for example be adapted according to the teachings disclosed in Hames B.D. and Higgins S.J. (1985) Nucleic Acid Hybridization: A Practical Approach. Hames and Higgins Ed., IRL Press, Oxford; and Current Protocols in Molecular Biology (supra).

"Percent homology" is used herein to refer to both nucleic acid sequences and amino acid sequences. Amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology". To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90% or 95% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the

corresponding position in the second sequence, then the molecules are homologous at that position. The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology=# of identical positions/total # of positions 100).

The comparison of sequences and determination of percent homology between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Karlin and Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264-68, modified as in Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-77. Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) J. Mol. Biol. 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to the sequences of the invention. BLAST protein searches can be performed with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to the polypeptide sequences of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Research 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The term "polypeptide" refers to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude post-translational modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl, acetyl, phosphate, amide, lipid, carboxyl, acyl or carbohydrate groups are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The term "protein" as used herein may be used synonymously with the term "polypeptide" or may refer to, in addition, a complex of two or more polypeptides which may be linked by bonds other than peptide bonds, for example, such polypeptides making up the protein may be linked by disulfide

bonds. The term "protein" may also comprehend a family of polypeptides having identical amino acid sequences but different post-translational modifications, particularly as may be added when such proteins are expressed in eukaryotic hosts.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NPP, or a biologically active fragment or homologue thereof is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of a protein according to the invention (e.g. NPP or a biologically active fragment or homologue thereof) in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of a protein according to the invention having less than about 30% (by dry weight) of protein other than the NPP, more preferably less than about 20% of protein other than the protein according to the invention, still more preferably less than about 10% of protein other than the protein according to the invention, and most preferably less than about 5% of protein other than the protein according to the invention. When the protein according to the invention or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NPP or a biologically active fragment or homologue thereof in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of a NPP having less than about 30% (by dry weight) of chemical precursors or non-NPP chemicals, more preferably less than about 20% chemical precursors or non-NPP chemical, still more preferably less than about 10% chemical precursors or non-NPP chemical, and most preferably less than about 5% chemical precursors or non-NPP chemical.

The term "recombinant polypeptide" is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide.

The term "Novel Plasma Polypeptide" or "NPP" refers to a polypeptide comprising the sequence selected from the group consisting of SEQ ID NOs:1-122. Such polypeptide may be post-translationally modified, for example, phosphorylated, acylated, or glycosylated. NPPs may also contain other structural or chemical modifications such as disulfide linkages or amino acid side chain

interactions such as hydrogen and amide bonds that result in complex secondary and tertiary structures. NPPs encompass functional signal sequences and mature and/or secreted amino acid species. NPPs also embrace mutant polypeptides, such as deletion, addition, swap, or truncation mutants, fusion polypeptides comprising such polypeptides, and polypeptide fragments of at least three, but preferably 6, 8, 10, 12, 15, or 21 contiguous amino acids of the sequence selected from the group consisting of SEQ ID NOs:1-122. The invention embodies polypeptides encoded by the nucleic acid sequences of the NPP-encoding genes or messenger RNAs, as well as the NPPs from humans, including isolated or purified NPPs consisting of, consisting essentially of, or comprising the sequence selected from the group consisting of SEQ ID NOs:1-122. Preferred NPPs retain at least one biological activity of NPP. Especially preferred are fragments of at least 6 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NOs:1-122 capable of binding or generating an antibody.

The term "biological activity" as used herein refers to any single function carried out by a NPP. These include but are not limited to: (1) indicating that an individual has or will have a cardiovascular disorder; (2) circulating through the bloodstream; (3) antigenicity, or the ability to bind an anti-NPP specific antibody; (4) immunogenicity, or the ability to generate an anti-NPP specific antibody; (5) interaction with a NPP target molecule; and (6) undergoing posttranslational processing, for example, specific proteolysis.

As used herein, a "NPP modulator" is a molecule (e.g., polynucleotide, polypeptide, small molecule, or antibody) that is capable of modulating (i.e., increasing or decreasing) either the expression or biological activity of the NPP of the invention. A NPP modulator that enhances NPP expression or activity is described as a NPP activator or agonist. Conversely, a NPP modulator that represses NPP expression or activity is described as a NPP inhibitor or antagonist. Preferably, NPP modulators increase/ decrease the expression or activity by at least 5, 10, or 20%. NPP inhibitors include anti-NPP antibodies, fragments thereof, antisense polynucleotides, and molecules characterized by screening assays, as described herein. NPP agonists include polynucleotide expression vectors and molecules characterized by screening assays as described herein.

A "NPP-related disorder" or "NPP-related disease" describes a cardiovascular disorder. Preferred disorders include coronary artery disease (CAD), coronary heart disease (CHD), peripheral vascular disease, cerebral ischemia (stroke), congestive heart failure, atherosclerosis, hypertension, and other cardiovascular diseases. The likelihood that a tested individual will develop or already has such a disorder is indicated by a difference in the plasma levels of at least one NPP between tested and control individuals.

Another aspect of the invention pertains to anti-NPP antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin

molecules, i.e., molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as a NPP or a biologically active fragment or homologue thereof. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind a NPP or a biologically active fragment or homologue thereof. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen-binding site capable of immunoreacting with a particular epitope of a NPP. A monoclonal antibody composition thus typically displays a single binding affinity for a particular NPP with which it immunoreacts.

As used herein, a "label group" is any compound that, when attached to a polynucleotide or polypeptide (including antibodies), allows for detection or purification of said polynucleotide or polypeptide. Label groups may be detected or purified directly or indirectly by a secondary compound, including an antibody specific for said label group. Useful label groups include radioisotopes (e.g., ^{32}P, ^{35}S, ^{3}H, ^{125}I), fluorescent compounds (e.g., 5-bromodesoxyuridin, umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride, phycoerythrin acetylaminofluorene, digoxigenin), luminescent compounds (e.g., luminol, GFP, luciferin, aequorin), enzymes or enzyme co-factor detectable labels (e.g., peroxidase, luciferase, alkaline phosphatase, galactosidase, or acetylcholinesterase), or compounds that are recognized by a secondary factor such as strepavidin, GST, or biotin. Preferably, a label group is attached to a polynucleotide or polypeptide in such a way as to not interfere with the biological activity of the polynucleotide or polypeptide.

Radioisotopes may be detected by direct counting of radioemission, film exposure, or by scintillation counting, for example. Enzymatic labels may be detected by determination of conversion of an appropriate substrate to product, usually causing a fluorescent reaction. Fluorescent and luminescent compounds and reactions may be detected by, e.g., radioemission, fluorescent microscopy, fluorescent activated cell sorting, or a luminometer.

As used herein with respect to antibodies, an antibody is said to "selectively bind" to a target if the antibody recognizes and binds the target of interest but does not substantially recognize and bind other molecules in a sample, e.g., a biological sample, which includes the target of interest.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome.

Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

As used herein, "effective amount" describes the amount of an agent, preferably a CPP modulator of the invention, sufficient to have a desired effect. For example, an anticardiovascular disorder effective amount is the amount of an agent required to reduce a symptom of a cardiovascular disorder in an individual by at least 1, 2, 5, 10, 15, or preferably 25%. The term may also describe the amount of an agent required to ameliorate a cardiovascular disorder-caused symptom in an individual. Common symptoms of cardiovascular disorders include: chest pressure, heartburn, nausea, vomiting, numbness, shortness of breath, heavy cold sweating, unexplained fatigue, and feelings of anxiety. The more severe symptoms of cardiovascular disorders are chest pain (angina pectoris), rhythm disturbances (arrhythmias), stroke, or heart attack. The effective amount for a particular patient may vary depending on such factors as the diagnostic method of the symptom being measured, the state of the condition being treated, the overall health of the patient, method of administration, and the severity of side-effects.

"HMMGene" describes a gene finder algorithm employed by the methods of the invention. HMMGene builds on a hidden Markov model (Durbin, R.M., et al., 1998, Biological sequence analysis, Cambridge University Press, Cambridge, UK) that recognizes: intergenic regions; 5' and 3' UTRs; coding regions; introns in both UTRs and coding regions; translation start and stop sites; splice sites; branchpoints; and poly(A) sites. The model is estimated by conditional maximum likelihood from training data (Krogh, A., 1997, "Two methods for improving performance of a HMM and their application for gene finding" Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology, p179-186. AAAI Press, Menlo Park, CA). Each state of the model is labeled as belonging to one of the nine classes: intergenic, 5' UTR, 3' UTR, coding, intron of phase 0, 1, or 2 in coding region, intron in 5' UTR, or intron in 3' UTR. Each path through the model gives a labeling of the DNA sequence. The total probability of a labeling is the sum over all paths giving that labeling.

Genes are predicted as the most probable labeling given the model by the *N*-best algorithm. This is an approximative algorithm, because there is no efficient way to do it exactly, but the approximation is very good (Krogh, A. 2000, Genome Research 10:523-328).

NPPs of the invention

The NPPs of the invention are described in the sequence listing as SEQ ID NOs:1-122. NPPs are encoded by genomic regions that have not been predicted to be coding sequences. The cDNA sequences encoding selected NPPs are described as SEQ ID NOs:123-138. Furthermore, NPPs are secreted in human plasma at differential levels in individuals that have or are at risk of developing a cardiovascular disorder (see Table 1). Thus, the NPPs of the invention provide not only novel human proteins and nucleotides with diagnostic utility, but provide the field of biotechnology with new information about the structure and potential of the human genome.

Preferred NPPs are polypeptides comprising an amino acid sequence selected from the group consisting of SEQ ID NOs:18, 20, 27, 30, 31, 43, 47, 53, 55, 62, 66, 67, 73, 76, 96, and 102. Preferably, such NPPs also comprise additional amino acids from one of the corresponding full length sequences of SEQ ID NOs:107-122 (see Table 1). Such additional amino acids are fused in frame with the selected sequence to form contiguous amino acid sequence from a polypeptide selected from the group consisting of SEQ ID NOs:107-122. Especially preferred peptide sequences include SEQ ID NOs:18, 20, 27, 31, 47, 53, 62, 66, 73, and 76, corresponding to the preferred full length proteins of SEQ ID NOs: 107-109, 111, 113, 114, 116, 117, 119, and 120. The polypeptide sequences of SEQ ID NOs:107-122 are encoded by the cDNAs described as SEQ ID NOs:123-138, in order.

One aspect of the invention pertains to isolated NPPs, biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-NPP antibodies. In one embodiment, native NPPs can be isolated from plasma, cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NPPs are produced by recombinant DNA techniques. Alternative to recombinant expression, a NPP peptides or polypeptides can be synthesized chemically using standard peptide synthesis techniques.

Typically, biologically active portions comprise a domain or motif with at least one activity of a NPP. A biologically active NPP may, for example, comprise at least 1, 2, 3, or 5 amino acid changes from the sequence selected from the group consisting of SEQ ID NOs:1-122, or comprise at least 1%, 2%, 3%, 5%, 8%, 10% or 15% changes in amino acids from the sequence selected from the group consisting of SEQ ID NOs: 1-122. In a preferred embodiment, a NPP comprises a target binding region and/or a signal sequence. The invention also concerns the polypeptide encoded by the NPP nucleotide sequences of the invention (SEQ ID NOs:123-138), or a complementary sequence thereof

or a fragment thereof.

In other embodiments, the NPP is substantially homologous to the sequence selected from the group consisting of SEQ ID NOs: 1-122, and retains the functional activity of the NPP, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described further herein. Accordingly, in another embodiment, the NPP is a protein which comprises an amino acid sequence which shares more than about 60% but less than 100% homology with the amino acid sequence selected from the group consisting of SEQ ID NOs:1-122 and retains the functional activity of the NPP selected from the group consisting of SEQ ID NOs:1-122. Preferably, the protein is at least about 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 92%, 95%, 97%, 98%, 99% or 99.8% homologous to the NPP selected from the group consisting of SEQ ID NOs:1-122, but is not identical to said sequence. Preferably the NPP is less than identical (e.g. 100% identity) to a naturally occurring NPP. Percent homology can be determined as further detailed herein.

Characterization of NPPs

The polypeptides of the invention, NPPs, are defined by the tryptic peptides of SEQ ID NOs:1-106 (Table 1). These peptides were isolated from the blood plasma from individuals with Coronary Artery Disease or healthy controls and characterized according to either the MicroProt.TM or MacroProt.TM method, as described in Examples 2 and 3, respectively. The peptides were identified using mass spectrometry and accompanying software, as described in Example 4. The tryptic peptides of the invention are encoded by genomic sequences that were previously characterized as noncoding regions. The identification program matches mass spectrometry data to amino acid sequences obtained by translating genomic sequences in all six frames. Genomic sequences from the NCBI (Genbank) database, 19 July 2001 version, were used. The Genbank accession number and translation frame are indicated in Table 1. For peptides translated in frames 1-3, the start and end positions of the coding sequence relative to the start of NCBI polynucleotide sequence are given. For peptides translated in frames 4-6, the start and end positions are given relative to the end of the indicated NCBI sequence.

In addition, some full length coding region were found using a HMMgene algorithm. SEQ ID NOs:107-122 represent these selected full length polypeptide sequences from which the corresponding tryptic peptides were released.

The Olav scores given in Table 1 reflect, among other things, the strength of the experimental MS-MS signal over noise, as detected by the MS-MS data identification software, and thus give an indication of the protein concentration in the sample. Where the peptide is found in both CAD and control plasma, the ratio of protein levels in CAD versus control plasma samples may calculated by one of many possible methods. One such method calculates the CAD/Control ratio by the number of

fractions from each sample containing the particular NPP (see Table 1). For example, the peptide of SEQ ID NO:106 is present in 1 CAD sample and 7 Control samples, indicating that the peptide is present at 0.14 times the normal level in CAD plasma. The peptide of SEQ ID NO:52, on the other hand, is found in 7 CAD samples and only 2 Controls, indicating a 3.5-fold increase in diseased plasma. Alternatively, and more accurately, the Olav scores obtained for each peptide in the mass spectrometry data analysis software are used to give a weighted ratio. For this method, the scores are added and the sum of the CAD scores is divided by the sum of the control scores for each peptide. Using the same examples, the peptide of SEQ ID NO:106 is present at 35.8/ 343.38, or 0.10, times the normal level in CAD plasma. The CAD/ Control ratio for the peptide of SEQ ID NO:52 is 234.84/ 47.4, or 4.95.

The methods of protein separation and identification according to the invention are extremely sensitive. The Microprot.TM process is able to detect very low abundance proteins with a plasma concentration in the range of a few hundreds of pM. Thus, the absence of a peptide in either control or disease plasma indicates that the particular NPP is present at a vanishingly low level in that plasma sample, if at all.

The coding regions corresponding to NCBI accession numbers NT_007091.5, NT_007897.5, NT_005986.5, NT_015926.5, and NT_026437.3 correspond to more than one peptide of the invention. Thus, the corresponding peptides which include SEQ ID NOs:21-22, 29-30, 14-15, 63-64, and 88-94, respectively, are preferred peptides of the invention.

As described in Examples 2 and 3, the plasma samples are subjected to a number of chromatography separations. Details about these chromatography methods are given in the Examples.

NPPs below 20kD were separated as described in Example 2. The first separation is on a cation exchange chromatography column, which is eluted with increasing salt concentration. Eighteen fractions are collected. The CEX column in Table 2 lists which fraction contained each tryptic peptide, as well as its elution conditions. Separation by cation exchange provides an indication of the overall positive charge of a polypeptide species. Cation exchange is followed by a reverse phase HPLC separation. The RP1 column in Table 2 lists in which of the 30 fractions each tryptic peptide eluted, as well as its elution conditions. Separation by reverse phase provides an indication of the overall hydrophobicity of a polypeptide species. The last two digits of the column labeled Run Number indicate which of the 24 eluted fractions from the reverse phase HPLC separation contained the peptide. NPPs above 20kD were separated as described in Example 3. Briefly, after initial labelling and depletion steps, the proteins are fractionated, among other things, by a benzamidine column, a red sepharose column, and a Rotofor® (Bio-Rad) apparatus, before being run on 2D gels. Full details are provided in Example 3 and in the corresponding Table 3.

TABLE 1

TABLE 1 Peptide SEQ ID NO (Protein, Nucleotide, if applicable)	Peptide Sequence	NCBI Acc. Number	Frame	Peptide Start	Peptide End	Proteome	Olav Score
1	SICPSALIKISLER	NT_004302.5	5	2263552	2263593	CAD	65.45
2	IASSGKTGIQTK	NT_004568.5	1	2681084	2681119	CAD	30.03
							33
							28.34
							25.67
							29.33
							13.44
							24.43
							25.14
3	HPNMLTECLLCGK	NT_004732.5	5	555876	555914	CAD	27.81
						CAD	35.13
						Control	31.39
						CAD	41.98
4	RLGHGIDAQ	NT_004745.5	3	877374	877400	CAD	38.05
							43.41
5	KMPLFIYICTK	NT_004836.5	6	3141499	3141531	CAD	13.4
							21.77
6	SAAHLILLR	NT_004893.5	3	788012	788038	CAD	23.85
						Control	28.82
						Control	14.33
						Control	28.9
						Control	29.67
						CAD	27.61
7	LPTTMLIGR	NT_005129.5	3	1166786	1166812	CAD	26.38
8	QMVLMSCVLK	NT_005244.5	6	314405	314434	Control	31.99
							36.79
9	QILENQVR	NT_005265.5	6	1003756	1003779	CAD	25.7
10	CVSSYPTSAEK	NT_005274.5	6	482937	482969	CAD	42.79
11	LCVLIMK	NT_005314.5	1	461085	461105	CAD	20.26
							23.37
12	STNAHLGAKR	NT_005466.1	6	182845	182874	CAD	16.54
						Control	32.02
13	FGKTDNINCPK	NT_005646.5	1	266303	266335	CAD	13.76
							17.84
14	WSPECSSTSIVLR	NT_005986.5	1	951326	951364	Control	38.4
15	GGNVCGTVANGKQEK	NT_005986.5	2	969039	969071	Control	46.57
16	HTNYFLKNHS	NT_006138.5	1	1424107	1424136	CAD	25.62
							37.62
17	MVLDVSDNEMTFSK	NT_006169.5	6	1016770	1016811	Control	27.52
18 (107, 123)	VMLMIQETNK	NT_006302.5	1	1224027	1224056	CAD	22.74
							44.83
							12.3
							35.85
							41.77

TABLE 1 Peptide SEQ ID NO (Protein, Nucleotide, if applicable)	Peptide Sequence	NCBI Acc. Number	Frame	Peptide Start	Peptide End	Proteome	Olav Score
							43.05
							45.15
							43.04
							45.86
19	IVHQVSKLFK	NT_006308.4	1	537523	537552	CAD	34.3
20 (108, 124)	LLNNFPYR	NT_006431.5	1	2363587	2363610	CAD	29.15
21	RPLSSSHIGSPR	NT_007091.5	3	648262	648297	CAD	27.74
22	KGAPLLGK	NT_007091.5	1	695582	695605	Control	4.46
23	RMNSAFGGR	NT_007096.5	5	211280	211306	Control	34.88
24	QGSGHIGK	NT_007116.3	4	100011	100034	Control	16.35
						CAD	17.91
						CAD	28.75
						Control	16.67
						Control	28.81
						Control	20.38
						Control	15.48
						CAD	10.07
						CAD	22.38
						CAD	17.83
						CAD	18.3
						CAD	21.93
						Control	20.4
						CAD	17.5
						CAD	11.41
						CAD	11.43
						CAD	19.37
						CAD	15.33
						CAD	22.79
						CAD	16.14
						CAD	16.67
						CAD	18.3
						CAD	19.36
						CAD	23.03
						Control	28.83
						CAD	17.4
						CAD	34.44
						CAD	25.58
						Control	13
						Control	13.42
						Control	13.52
25	KAVNALAHK	NT_007592.5	4	9987357	9987383	CAD	51.13
26	LIFVCEASLHPK	NT_007592.5	5	2604706	2604741	CAD	33.01
27 (109, 125)	SGCTNLRSHQQCIR	NT_007712.5	3	586740	586781	CAD	59.17
28	QGWQGNSIGKK	NT_007793.5	1	1383447	1383479	Control	48.87
29	LVPVLQI	NT_007897.5	5	2418161	2418181	CAD	23.06
30 (110, 126)	TEGLTLLQLV	NT_007897.5	2	1266504	1266533	CAD	11.07

TABLE 1 Peptide SEQ ID NO (Protein, Nucleotide, if applicable)	Peptide Sequence	NCBI Acc. Number	Frame	Peptide Start	Peptide End	Proteome	Olav Score
31 (111, 127)	ESIYFIIAAMLVATK	NT_007914.5	1	1375030	1375074	CAD	44.99
32	IFLLGQITSIPDKL	NT_007930.1	5	1340892	1340933	Control	38.16
33	KPLKNGSQFS	NT_008117.5	5	757179	757208	Control	50.01
34	RVITPLIK	NT_008421.5	3	3847592	3847615	CAD	11.23
35	LGTVSLTH	NT_008476.5	2	1908502	1908525	CAD	30.96
36	RHCLLFVCFCK	NT_008541.5	5	609972	610004	CAD	26.95
37	CHFCLTCSR	NT_008609.5	5	3864355	3864381	CAD	25.39
38	IPTTFETNL	NT_008669.5	5	687617	687643	CAD	39.53
39	STVLSASLHLR	NT_008682.5	5	1820942	1820974	Control	47.73
							32.84
40	LEVELTFLWPSPPR	NT_008682.5	1	1935267	1935308	CAD	41.57
41	IFLTMDQLLQN	NT_008984.5	6	7627020	7627052	CAD	24.82
42	SASLMEIQSKK	NT_009276.5	4	954299	954331	CAD	51.07
43 (112, 128)	MKPLVDYK	NT_009561.5	1	833003	833026	CAD	29.91
44	EDLGSKGPK	NT_009678.5	6	671983	672009	CAD	31.17
45	CLLLRGHYSAMR	NT_009700.1	5	1390503	1390538	Control	24.63
46	TFSSALFWK	NT_009799.1	5	5436983	5437009	CAD	30.18
47 (113, 129)	QADGTVFSK	NT_009891.1	3	1001736	1001762	Control	31.27
48	SILFAFSLYR	NT_009952.5	5	4845761	4845790	CAD	22.75
						CAD	39.32
						Control	33.11
						CAD	26.38
						CAD	36.39
						CAD	29.98
						CAD	29.64
49	SSPLDLVCNSSSTSY	NT_009967.5	5	112490	112534	CAD	30.62
50	VKMLHALVLK	NT_010289.5	5	3644894	3644923	CAD	46
51	ADSGLAQSDGK	NT_010558.5	5	2417	2449	CAD	23.13
52	DHEDAWRMFSAR	NT_010771.5	4	826322	826357	CAD	40.77
						Control	18.78
						Control	28.62
						CAD	27.25
						CAD	36.04
						CAD	43.65
						CAD	41.26
						CAD	33.64
						CAD	32.34
53 (114, 130)	CVIFPLNSYGMLLK	NT_010909.5	3	184868	184909	CAD	20.66
						CAD	36.34
						Control	46.24
54	HLKLAISSLLR	NT_010966.5	3	2333085	2333117	CAD	58.74
55 (115, 131)	DSYLNVKR	NT_011387.5	2	7014161	7014184	CAD	39.58
						Control	13.82
							27.08
							30.28
							34.27

TABLE 1

Peptide SEQ ID NO (Protein, Nucleotide, if applicable)	Peptide Sequence	NCBI Acc. Number	Frame	Peptide Start	Peptide End	Proteome	Olav Score
56	HSELCLAR	NT_011387.5	3	7008231	7008254	CAD	11.88
57	CSKTFINTK	NT_011522.3	4	623590	623616	CAD	29.95
						Control	19.3
58	NRQTLLLLMSCR	NT_011568.5	5	424837	424872	CAD	21.54
59	YLSDGWIKGYIK	NT_011588.5	5	886577	886612	CAD	23.33
60	DVSSAIPNSVS	NT_011834.3	2	84883	84911	Control	42.19
61	VSWHKHLLLLR	NT_011875.6	3	1869148	1869180	Control	37.06
62 (116, 132)	EAEFESTMQK	NT_011896.6	1	1580932	1580961	Control	52
63	ILMDDFKK	NT_015926.5	5	1755932	1755955	CAD	24.47
64	YSEIKEK	NT_015926.5	1	465732	465752	Control	12.54
65	SRHQEIGCLAR	NT_017582.5	4	541155	541187	Control	39.72
66 (117, 133)	QVQSYHVLGK	NT_019265.5	1	749486	749515	Control	39.81
67 (118, 134)	NPMKIFEK	NT_019546.5	1	611483	611506	Control	30.52
68	LPNHLLNHR	NT_019599.5	5	852111	852137	Control	27.44
69	ETLMAAELNMAGIYNGIKGAR	NT_021877.5	5	1739489	1739551	CAD	29.72
						CAD	71.58
						Control	56.2
						Control	71.59
						Control	58.66
70	PLTLWSHR	NT_021907.5	5	981290	981313	CAD	35.57
71	SSLVLYVLR	NT_021942.3	5	67855	67881	Control	24.96
72	MPGILYNK	NT_022136.5	5	202135	202158	Control	30.34
73 (119, 135)	CLCTHNGASKYMK	NT_022148.5	1	530381	530419	CAD	43.19
							34.4
							24.59
							28.51
74	LGFLFVSETESR	NT_022443.5	4	797427	797462	Control	36.28
75	ICNIQQAHIHWR	NT_022762.4	5	102375	102410	Control	33.85
76 (120, 136)	EQNKILSNLEIER	NT_022851.5	3	187441	187479	CAD	12.97
77	CLYSFVFSR	NT_022938.3	5	64842	64868	Control	38.96
							37.17
78	ENVIPSLTVPK	NT_023195.5	6	2027050	2027082	CAD	37.92
79	KTILEHIPLR	NT_023399.5	5	420089	420118	CAD	28.13
80	KSCVGLTTFY	NT_023929.5	2	79146	79175	CAD	37.63
81	LSAAVRLSAAVR	NT_023957.5	6	1313220	1313255	CAD	38.71
82	QQHKSASLLR	NT_024037.5	3	1340003	1340032	CAD	21.83
83	QDHLNISYK	NT_024653.5	1	68486	68512	Control	46.67
84	INEKIFCGHK	NT_025682.2	2	965	994	CAD	26.37
85	CTSVDHTPIR	NT_025741.4	6	1773833	1773862	CAD	28.36
86	SHLNVQSEKVK	NT_026231.1	6	192651	192683	CAD	34.27
87	YALKCHNLQILHTK	NT_026302.3	6	522507	522548	CAD	13.93
88	FCKFSLLISSSTR	NT_026437.3	6	36420692	36420730	CAD	33.22

TABLE 1 Peptide SEQ ID NO (Protein, Nucleotide, if applicable)	Peptide Sequence	NCBI Acc. Number	Frame	Peptide Start	Peptide End	Proteome	Olav Score
89	FSDDTHRTGR	NT_026437.3	4	21808437	21808466	Control	27.16
						CAD	37.31
90	TAVVSLPR	NT_026437.3	3	23686612	23686635	Control	32.23
						CAD	39.89
91	EQLSLLDR	NT_026437.3	6	14621377	14621400	Control	34
						Control	24.29
92	AVLDVFEEGTEASAATAVK	NT_026437.3	4	14630126	14630182	CAD	16.94
93	ITLLSALVETR	NT_026437.3	4	14630183	14630215	Control	29.54
						Control	13.9
94	ILHMLCHLILIR	NT_026437.3	4	5302542	5302577	CAD	33.51
95	IHQQLALWTWK	NT_027054.2	4	100514	100546	Control	36.61
96 (121, 137)	PEMVVQACSLSY	NT_027064.2	3	578258	578293	CAD	43.05
97	LMYLVFTKASPK	NT_027193.2	3	80270	80305	CAD	45.27
98	EDNTAEYEPCALR	NT_028089.1	6	46146	46184	Control	34.23
99	WFLRILGSPMGVLSQWGK	NT_028225.2	6	156449	156502	CAD	24.18
100	GTELLIHHQWPK	NT_028360.2	5	1069868	1069903	CAD	23.21
101	ALHLDNSAFR	NT_028389.1	1	115143	115172	CAD	44.13
102 (122, 138)	NAKISQAPW	NT_028428.2	1	296117	296117	Control	23.36
103	CWATESNEIHLEIQT	NT_029250.1	4	69320	69364	CAD	29
104	LFLDCMLNK	NT_029315.1	4	941255	941281	CAD	40.37
105	LFIFTCVFHK	NT_029331.1	3	301531	301560	Control	25.54
106	HCRTNHVLLLLR	NT_029391.1	1	992052	992087	Control	25.33
						CAD	35.8
						Control	62.26
						Control	48.59
						Control	58.33
						Control	49.57
						Control	49.59
						Control	59.72
						Control	15.32

NPP nucleic acids

One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode NPPs or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. The polynucleotides of the invention represent novel human coding sequences and thus reveal new potential of the human genome to the field of biotechnology. Some of the polynucleotides of the invention were characterized by applying a HMMgene prediction to the NPPs found in human plasma samples (see "Characterization of NPPs"). Said nucleic acids may be used for gene mapping, protein expression, and diagnostic methods as further described herein.

An object of the invention is a purified, isolated, or recombinant nucleic acid selected from the group consisting of SEQ ID NOs:123-138. The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity with a polynucleotide selected from the group consisting of SEQ ID NOs:123-138, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide selected from the group consisting of SEQ ID NOs:123-138, or a sequence complementary thereto or a biologically active fragment thereof.

Another object of the invention is a purified, isolated, or recombinant nucleic acid coding for a NPP selected from the group consisting of SEQ ID NOs:1-122, complementary sequences thereto, and fragments thereof. The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity with a polynucleotide coding for a NPP, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide coding for a NPP, or a sequence complementary thereto or a biologically active fragment thereof. Another object of the invention relates to purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide coding for a NPP, or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof.

Purified or recombinant nucleic acids coding for a NPP selected from the group consisting of SEQ ID NOs:1-122 can be obtained by using the information provided in Table 1, to localize the NCBI entry containing the DNA of interest, and within this entry to localize the precise region of interest (Table 1 discloses the positions, in bp, of the DNA sequence coding for the tryptic peptide within the NCBI entry).

Degenerate polynucleotides, primers, and probes may be designed using a protein sequence selected from the group consisting of SEQ ID NOs:1-106. Degenerate polynucleotide sequences are useful for cloning full length NPP-encoding cDNAs and related sequences. For example, degenerate PCR primers may be used to amplify portions of the coding sequence for a protein. Amplified fragments may then be sequenced and put in order by matching sequence ends or used to generate additional primers, if necessary. Such cloning techniques are described by Piraee and Vining (J Industrial Microbiology & Biotechnology, 2002, 29:1-5). Degenerate sequences may be designed using common algorithms (e.g., CODEHOP March 2003 version, Rose, et al., Nucleic Acids Research, 1998, 26:1628-1635). Amplified sequences may then be cloned using methods common to the art (for example, those in Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning, A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

and Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1994) Current Protocols in Molecular Biology. Wiley, New York).

In another preferred aspect, the invention pertains to purified or isolated nucleic acid molecules that encode a portion or variant of a NPP, wherein the portion or variant displays a NPP biological activity of the invention. Preferably said portion or variant is a portion or variant of a naturally occuring full-length NPP.

The nucleotide sequence determined from the previously uncharacterized NPP gene allows for the generation of probes and primers designed for use in identifying and/or cloning NPP cDNAs or other NPP family members (e.g. sharing the novel functional domains), as well as NPP homologues from other species.

A nucleic acid fragment encoding a "biologically active portion of a NPP" can be prepared by isolating a portion of a nucleotide sequence coding for a NPP, which encodes a polypeptide having a NPP biological activity, expressing the encoded portion of the NPP (e.g., by recombinant expression in vitro or in vivo) and assessing the activity of the encoded portion of the NPP.

The invention further encompasses nucleic acid molecules that differ from the NPP nucleotide sequences of the invention due to degeneracy of the genetic code and encode the same NPPs and fragments of the invention.

In addition to the NPP nucleotide sequences described above, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NPPs may exist within a population (e.g., the human population). Such genetic polymorphism may exist among individuals within a population due to natural allelic variation. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NPP gene or nucleic acid sequence encoding the NPP. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NPP nucleic acids of the invention can be isolated based on their homology to the NPP nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

It will be appreciated that the invention comprises polypeptides having an amino acid sequence encoded by any of the polynucleotides of the invention.

Uses of NPP nucleic acids

Polynucleotide sequences (or the complements thereof) encoding NPPs have various applications, including uses as hybridization probes, in chromosome mapping and gene cloning, and for the preparation of NPPs by recombinant techniques, as described herein. The polynucleotides

described herein, including sequence variants thereof, can be used in diagnostic assays. Accordingly, diagnostic methods based on detecting the presence of such polynucleotides in body fluids or tissue samples are a feature of the present invention. Examples of nucleic acid based diagnostic assays in accordance with the present invention include, but are not limited to, hybridization assays, e.g., in situ hybridization, and PCR-based assays. Polynucleotides, including extended length polynucleotides, sequence variants and fragments thereof, as described herein, may be used to generate hybridization probes or PCR primers for use in such assays. Such probes and primers will be capable of detecting polynucleotide sequences, including genomic sequences that are similar, or complementary to, the NPP polynucleotides described herein.

The invention includes primer pairs for carrying out a PCR to amplify a segment of a polynucleotide of the invention. Each primer of a pair is an oligonucleotide having a length of between 15 and 30 nucleotides such that i) one primer of the pair forms a perfectly matched duplex with one strand of a polynucleotide of the invention and the other primer of the pair form a perfectly match duplex with the complementary strand of the same polynucleotide, and ii) the primers of a pair form such perfectly matched duplexes at sites on the polynucleotide that separated by a distance of between 10 and 2500 nucleotides. Preferably, the annealing temperature of each primer of a pair to its respective complementary sequence is substantially the same.

Hybridization probes derived from polynucleotides of the invention can be used, for example, in performing in situ hybridization on tissue samples, such as fixed or frozen tissue sections prepared on microscopic slides or suspended cells. Briefly, a labeled DNA or RNA probe is allowed to bind its DNA or RNA target sample in the tissue section on a prepared microscopic, under controlled conditions. Generally, dsDNA probes consisting of the DNA of interest cloned into a plasmid or bacteriophage DNA vector are used for this purpose, although ssDNA or ssRNA probes may also be used. Probes are generally oligonucleotides between about 15 and 40 nucleotides in length. Alternatively, the probes can be polynucleotide probes generated by PCR random priming primer extension or in vitro transcription of RNA from plasmids (riboprobes). These latter probes are typically several hundred base pairs in length. The probes can be labeled by any of a number of label groups and the particular detection method will correspond to the type of label utilized on the probe (e.g., autoradiography, X-ray detection, fluorescent or visual microscopic analysis, as appropriate). The reaction can be further amplified in situ using immunocytochemical techniques directed against the label of the detector molecule used, such as an antibody directed to a fluorescein moiety present on a fluorescently labeled probe. Specific labeling and in situ detection methods can be found, for example, in Howard, G. C., Ed., *Methods in Nonradioactive Detection,* Appleton & Lange, Norwalk, Conn., (1993).

Hybridization probes and PCR primers may also be selected from the genomic sequences

corresponding to the full-length proteins identified in accordance with the present invention, including promoter, enhancer elements and introns of the gene encoding the naturally occurring polypeptide. Nucleotide sequences encoding a NPP can also be used to construct hybridization probes for mapping the gene encoding NPPs and for the genetic analysis of individuals. Individuals carrying variations of, or mutations in the NPP gene may be detected at the DNA level by a variety of techniques. Nucleic acids used for diagnosis may be obtained from a patient's cells, including, for example, tissue biopsy and autopsy material. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki, et al. *Nature* 324:163-166 (1986)) prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid of the present invention can be used to identify and analyze mutations in the gene of the present invention. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabeled RNA of the invention or alternatively, radiolabeled antisense DNA sequences of the invention. Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and Sl protection or the chemical cleavage method (e.g. Cotton, et al., *Proc. Natl. Acad. Sci. USA* 85:4397-4401 (1985)), or by differences in melting temperatures. "Molecular beacons" (Kostrikis L. G. et al., *Science* 279:1228-1229 (1998)), hairpin-shaped, single-stranded synthetic oligonucleotides containing probe sequences which are complementary to the nucleic acid of the present invention, may also be used to detect point mutations or other sequence changes as well as monitor expression levels of NPPs.

Oligonucleotide and Antisense Compounds

Oligonucleotides of the invention, including PCR primers and antisense compounds, are synthesized by conventional means on a commercially available automated DNA synthesizer, e.g. an Applied Biosystems (Foster City, CA) model 380B, 392 or 394 DNA/RNA synthesizer, or like instrument. Preferably, phosphoramidite chemistry is employed, e.g. as disclosed in the following references: Beaucage and Iyer, Tetrahedron, 48: 2223-2311 (1992); Molko et al, U.S. patent 4,980,460; Koster et al, U.S. patent 4,725,677; Caruthers et al, U.S. patents 4,415,732; 4,458,066; and 4,973,679; and the like. For therapeutic use, nuclease resistant backbones are preferred. Many types of modified oligonucleotides are available that confer nuclease resistance, e.g. phosphorothioate, phosphorodithioate, phosphoramidate, or the like, described in many references, e.g. phosphorothioates: Stec et al, U.S. patent 5,151,510; Hirschbein, U.S. patent 5,166,387; Bergot, U.S. patent 5,183,885; phosphoramidates: Froehler et al, International application PCT/US90/03138; and for a review of additional applicable chemistries: Uhlmann and Peyman (cited above). The length of the antisense oligonucleotides has to be sufficiently large to ensure that specific binding will take

place only at the desired target polynucleotide and not at other fortuitous sites. The upper range of the length is determined by several factors, including the inconvenience and expense of synthesizing and purifying oligomers greater than about 30-40 nucleotides in length, the greater tolerance of longer oligonucleotides for mismatches than shorter oligonucleotides, and the like. Preferably, the antisense oligonucleotides of the invention have lengths in the range of about 15 to 40 nucleotides. More preferably, the oligonucleotide moieties have lengths in the range of about 18 to 25 nucleotides.

Primers and probes

Primers and probes of the invention can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang SA et al (Methods Enzymol 1979;68:90-98), the phosphodiester method of Brown EL et al (Methods Enzymol 1979;68:109-151), the diethylphosphoramidite method of Beaucage et al (Tetrahedron Lett 1981, 22: 1859-1862) and the solid support method described in EP 0 707 592, the disclosures of which are incorporated herein by reference in their entireties.

Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in WO 92/20702, morpholino analogs which are described in U.S. Patents 5,185,444; 5,034,506 and 5,142,047. If desired, the probe may be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group.

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating any label group known in the art to be detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. Additional examples include non-radioactive labeling of nucleic acid fragments as described in Urdea et al. (Nucleic Acids Research. 11:4937-4957, 1988) or Sanchez-Pescador et al. (J. Clin. Microbiol. 26(10):1934-1938, 1988). In addition, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea et al (Nucleic Acids Symp. Ser. 24:197-200, 1991) or in the European patent No. EP 0225807 (Chiron).

A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which forms a binding pair

with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA labeling techniques are well known to the skilled technician.

The probes of the present invention are useful for a number of purposes. They can be notably used in Southern hybridization to genomic DNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in NPP-encoding genes or mRNA using other techniques.

Any of the nucleic acids, polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member attached to the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other

suitable animal's) red blood cells, duracytes and other configurations known to those of ordinary skill in the art. The nucleic acids, polynucleotides, primers and probes of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on a solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips® (Affymetrix, Santa Clara, CA), and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092.

Methods for obtaining variant nucleic acids and polypeptides

In addition to naturally-occurring allelic variants of the NPP sequences that may exist in the population, the skilled artisan will appreciate that changes can be introduced by mutation into the nucleotide sequences coding for NPPs, thereby leading to changes in the amino acid sequence of the encoded NPP, with or without altering the functional activity.

Several types of variants are contemplated including 1) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue and such substituted amino acid residue may or may not be one encoded by the genetic code, or 2) one in which one or more of the amino acid residues includes a substituent group, or 3) one in which the mutated NPP is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or 4) one in which the additional amino acids are fused to the NPP, such as a leader, a signal or anchor sequence, a sequence which is employed for purification of the NPP, or sequence from a precursor protein. Such variants are deemed to be within the scope of those skilled in the art.

For example, nucleotide substitutions leading to amino acid substitutions can be made in the

sequences that do not substantially change the biological activity of the protein. An amino acid residue-can be altered from the wild-type sequence encoding a NPP, or a biologically active fragment or homologue thereof without altering the biological activity. In general, amino acid residues that are shared among NPP homologues are predicted to be less amenable to alteration.

In another aspect, the invention pertains to nucleic acid molecules encoding NPPs that contain changes in amino acid residues that result in a modified biological activity. In another aspect, the invention pertains to nucleic acid molecules encoding NPPs that contain changes in amino acid residues that are essential for a NPP biological activity. Such NPPs differ in amino acid sequence from the sequence selected from the group consisting of SEQ ID NOs:1-122 and display reduced activity, or essentially lack one or more NPP biological activities.

Mutations, substitutions, additions, or deletions can be introduced into any of SEQ ID NOs:1-122, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. For example, conservative amino acid substitutions may be made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a NPP, or a biologically active fragment or homologue thereof may be replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NPP coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened biological activity to identify mutants that retain activity. Following mutagenesis of the nucleotide encoding one of SEQ ID NOs:1-122, the encoded protein can be expressed recombinantly and the activity of the protein can be determined in any suitable assay, for example, as provided herein.

The invention also provides NPP chimeric or fusion proteins. As used herein, a NPP "chimeric protein" or "fusion protein" comprises a NPP of the invention or fragment therof, operatively linked or fused in frame to a non-NPP sequence. In a preferred embodiment, a NPP fusion protein comprises at least one biologically active portion of a NPP. In another preferred embodiment, a NPP fusion protein comprises at least two biologically active portions of a NPP. For example, in one embodiment, the fusion protein is a GST- NPP fusion protein in which NPP domain sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification

of recombinant NPP. In another embodiment, the fusion protein is a NPP containing a heterologous signal sequence at its N-terminus, for example, to allow for a desired cellular localization in a certain host cell. In yet another embodiment, the fusion is a NPP biologically active fragment and an immunoglobulin molecule. Such fusion proteins are useful, for example, to increase the valency of NPP binding sites. The NPP fusion proteins of the invention can be used as immunogens to produce anti-NPP antibodies in a subject, to purify NPP or NPP ligands.

Chemical Manufacture of NPP Compositions

Peptides of the invention are synthesized by standard techniques (e.g. Stewart and Young, Solid Phase Peptide Synthesis, 2nd Ed., Pierce Chemical Company, Rockford, IL, 1984). Preferably, a commercial peptide synthesizer is used, e.g. Applied Biosystems, Inc. (Foster City, CA) model 430A, and polypeptides of the invention may be assembled from multiple, separately synthesized and purified, peptide in a convergent synthesis approach, e.g. Kent et al, U.S. patent 6,184,344 and Dawson and Kent, Annu. Rev. Biochem., 69: 923-960 (2000). Peptides of the invention may be assembled by solid phase synthesis on a cross-linked polystyrene support starting from the carboxyl terminal residue and adding amino acids in a stepwise fashion until the entire peptide has been formed. The following references are guides to the chemistry employed during synthesis: Schnolzer et al, Int. J. Peptide Protein Res., 40: 180-193 (1992); Merrifield, J. Amer. Chem. Soc., Vol. 85, pg. 2149 (1963); Kent et al., pg 185, in Peptides 1984, Ragnarsson, Ed. (Almquist and Weksell, Stockholm, 1984); Kent et al., pg. 217 in Peptide Chemistry 84, Izumiya, Ed. (Protein Research Foundation, B.H. Osaka, 1985); Merrifield, Science, Vol. 232, pgs. 341-347 (1986); Kent, Ann. Rev. Biochem., Vol. 57, pgs. 957-989 (1988), and references cited in these latter two references.

Preferably, chemical synthesis of polypeptides of the invention is carried out by the assembly of peptide fragments by native chemical ligation, as described by Dawson et al, Science, 266: 776-779 (1994) and Kent el al, U.S. patent 6,184,344. Briefly, in the approach a first peptide fragment is provided with an N-terminal cysteine having an unoxidized sulfhydryl side chain, and a second peptide fragment is provided with a C-terminal thioester. The unoxidized sulfhydryl side chain of the N-terminal cysteine is then condensed with the C-terminal thioester to produce an intermediate peptide fragment which links the first and second peptide fragments with a β-aminothioester bond. The β-aminothioester bond of the intermediate peptide fragment then undergoes an intramolecular rearrangement to produce the peptide fragment product which links the first and second peptide fragments with an amide bond. Preferably, the N-terminal cysteine of the internal fragments is protected from undesired cyclization and/ or concatenation reactions by a cyclic thiazolidine protecting group as described below. Preferably, such cyclic thiazolidine protecting group is a thioprolinyl group.

Peptide fragments having a C-terminal thioester may be produced as described in the following references: Kent et al, U.S. patent 6,184,344; Tam et al, Proc. Natl. Acad. Sci., 92: 12485-12489 (1995); Blake, Int. J. Peptide Protein Res., 17: 273 (1981); Canne et al, Tetrahedron Letters, 36: 1217-1220 (1995); Hackeng et al, Proc. Natl. Acad. Sci., 94: 7845-7850 (1997); or Hackeng et al, Proc. Natl. Acad. Sci., 96: 10068-10073 (1999). Preferably, the method described by Hackeng et al (1999) is employed. Briefly, peptide fragments are synthesized on a solid phase support (described below) typically on a 0.25 mmol scale by using the in situ neutralization/HBTU activation procedure for Boc chemistry disclosed by Schnolzer et al, Int. J. Peptide Protein Res., 40: 180-193 (1992). (HBTU is 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate and Boc is tert-butoxycarbonyl). Each synthetic cycle consists of N^{α}-Boc removal by a 1- to 2- minute treatment with neat TFA, a 1-minute DMF flow wash, a 10- to 20-minute coupling time with 1.0 mmol of preactivated Boc-amino acid in the presence of DIEA, and a second DMF flow wash. (TFA is trifluoroacetic acid, DMF is N,N-dimethylformamide, and DIEA is N,N-diisopropylethylamine). N^{α}-Boc-amino acids (1.1 mmol) are preactivated for 3 minutes with 1.0 mmol of HBTU (0.5 M in DMF) in the presence of excess DIEA (3 mmol). After each coupling step, yields are determined by measuring residual free amine with a conventional quantitative ninhydrin assay, e.g. as disclosed in Sarin et al, Anal. Biochem., 117: 147-157 (1981). After coupling of Gln residues, a DCM flow wash is used before and after deprotection by using TFA, to prevent possible high-temperature (TFA/DMF)-catalyzed pyrrolidone formation. After chain assembly is completed, the peptide fragments are deprotected and cleaved from the resin by treatment with anhydrous HF for 1 hour at 0°C with 4% *p*-cresol as a scavenger. The imidazole side-chain 2,4-dinitrophenyl (dnp) protecting groups remain on the His residues because the dnp-removal procedure is incompatible with C-terminal thioester groups. However, dnp is gradually removed by thiols during the ligation reaction. After cleavage, peptide fragments are precipitated with ice-cold diethylether, dissolved in aqueous acetonitrile, and lyophilized.

Thioester peptide fragments described above are preferably synthesized on a trityl-associated mercaptopropionic acid-leucine (TAMPAL) resin, made as disclosed by Hackeng et al (1999), or comparable protocol. Briefly, N^{α}-Boc-Leu (4 mmol) is activated with 3.6 mmol of HBTU in the presence of 6 mmol of DIEA and coupled for 16 minutes to 2 mmol of p-methylbenzhydrylamine (MBHA) resin, or the equivalent. Next, 3 mmol of S-trityl mercaptopropionic acid is activated with 2.7 mmol of HBTU in the presence of 6 mmol of DIEA and coupled for 16 minutes to Leu-MBHA resin. The resulting TAMPAL resin can be used as a starting resin for polypeptide-chain assembly after removal of the trityl protecting group with two 1-minute treatments with 3.5% triisopropylsilane and 2.5% H_2O in TFA. The thioester bond can be formed with any desired amino acid by using

standard in situ-neutralization peptide coupling protocols for 1 hour, as disclosed in Schnolzer et al (cited above). Treatment of the final peptide fragment with anhydrous HF yields the C-terminal activated mercaptopropionic acid-leucine (MPAL) thioester peptide fragments.

Preferably, thiazolidine-protected thioester peptide fragment intermediates are used in native chemical ligation under conditions as described by Hackeng et al (1999), or like conditions. Briefly, 0.1 M phosphate buffer (pH 8.5) containing 6 M guanidine, 4% (vol/vol) benzylmercaptan, and 4% (vol/vol) thiophenol is added to dry peptides to be ligated, to give a final peptide concentration of 1-3 mM at about pH 7, lowered because of the addition of thiols and TFA from the lyophilized peptide. Preferably, the ligation reaction is performed in a heating block at 37°C and is periodically vortexed to equilibrate the thiol additives. The reaction may be monitored for degree of completion by MALDI-MS or HPLC and electrospray ionization MS.

After a native chemical ligation reaction is completed or stopped, the N-terminal thiazolidine ring of the product is opened by treatment with a cysteine deprotecting agent, such as O-methylhydroxylamine (0.5 M) at pH 3.5-4.5 for 2 hours at 37° C, after which a 10-fold excess of Tris-(2-carboxyethyl)-phosphine is added to the reaction mixture to completely reduce any oxidizing reaction constituents prior to purification of the product by conventional preparative HPLC. Preferably, fractions containing the ligation product are identified by electrospray MS, are pooled, and lyophilized.

After the synthesis is completed and the final product purified, the final polypeptide product may be refolded by conventional techniques, e.g. Creighton, Meth. Enzymol., 107: 305-329 (1984); White, Meth. Enzymol., 11: 481-484 (1967); Wetlaufer, Meth. Enzymol., 107: 301-304 (1984). Preferably, a final product is refolded by air oxidation by the following: the reduced lyophilized product is dissolved (at about 0.1 mg/mL) in 1 M guanidine hydrochloride (or like chaotropic agent) with 100 mM Tris, 10 mM methionine, at pH 8.6. After gentle overnight stirring, the re-folded product is isolated by reverse phase HPLC with conventional protocols.

Recombinant Expression Vectors and Host Cells

The polynucleotide sequences described herein can be used in recombinant DNA molecules that direct the expression of the corresponding polypeptides in appropriate host cells. Because of the degeneracy in the genetic code, other DNA sequences may encode the equivalent amino acid sequence, and may be used to clone and express the NPPs. Codons preferred by a particular host cell may be selected and substituted into the naturally occurring nucleotide sequences, to increase the rate and/or efficiency of expression. The nucleic acid (e.g., cDNA or genomic DNA) encoding the desired NPP may be inserted into a replicable vector for cloning (amplification of the DNA), or for expression. The polypeptide can be expressed recombinantly in any of a number of expression

systems according to methods known in the art (Ausubel, et al., editors, Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1990). Appropriate host cells include yeast, bacteria, archebacteria, fungi, and insect and animal cells, including mammalian cells, for example primary cells, including stem cells, including, but not limited to bone marrow stem cells. More specifically, these include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors, and yeast transformed with yeast expression vectors. Also included, are insect cells infected with a recombinant insect virus (such as baculovirus), and mammalian expression systems. The nucleic acid sequence to be expressed may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site using techniques known in the art. Vector components generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

The NPPs of the present invention are produced by culturing a host cell transformed with an expression vector containing a nucleic acid encoding a NPP, under the appropriate conditions to induce or cause expression of the protein. The conditions appropriate for NPP expression will vary with the choice of the expression vector and the host cell, as ascertained by one skilled in the art. For example, the use of constitutive promoters in the expression vector may require routine optimization of host cell growth and proliferation, while the use of an inducible promoter requires the appropriate growth conditions for induction. In addition, in some embodiments, the timing of the harvest is important. For example, the baculoviral systems used in insect cell expression are lytic viruses, and thus harvest time selection can be crucial for product yield.

A host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the protein include, but are not limited to, glycosyl, acetyl, phosphate, amide, lipid, carboxyl, acyl, or carbohydrate groups. Post-translational processing, which cleaves a "prepro" form of the protein, may also be important for correct insertion, folding and/or function. By way of example, host cells such as CHO, HeLa, BHK, MDCK, 293, W138, etc. have specific cellular machinery and characteristic mechanisms for such post-translational activities and may be chosen to ensure the correct modification and processing of the introduced, foreign protein. Of particular interest are *Drosophila melanogaster* cells, *Saccharomyces cerevisiae* and other yeasts, *E. coli, Bacillus subtilis,* SF9 cells, Cl29 cells, 293 cells, Neurospora, BHK, CHO, COS, and HeLa cells, fibroblasts, Schwanoma cell lines, immortalized mammalian myeloid and lymphoid cell lines, Jukat cells, human cells and other primary cells.

The nucleic acid encoding a NPP must be "operably linked" by placing it into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" DNA sequences are contiguous, and, in the case of a secretory leader or other polypeptide sequence, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice. Promoter sequences encode either constitutive or inducible promoters. The promoters may be either naturally occurring promoters or hybrid promoters. Hybrid promoters, which combine elements of more than one promoter, are also known in the art, and are useful in the present invention. The expression vector may comprise additional elements, for example, the expression vector may have two replication systems, thus allowing it to be maintained in two organisms, for example in mammalian or insect cells for expression and in a procaryotic host for cloning and amplification. Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2: plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells. Further, for integrating expression vectors, the expression vector contains at least one sequence homologous to the host cell genome, and preferably, two homologous sequences which flank the expression construct. The integrating vector may be directed to a specific locus in the host cell by selecting the appropriate homologous sequence for inclusion in the vector. Constructs for integrating vectors are well known in the art. In an additional embodiment, a heterologous expression control element may be operably linked with the endogenous gene in the host cell by homologous recombination (described in US Patents 6410266 and 6361972). This technique allows one to regulate expression to a desired level with a chosen control element while ensuring proper processing and modification of NPP endogenously expressed by the host cell. Useful heterologous expression control elements include but are not limited to CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous Sarcoma Virus (RSV), and metallothionein promoters.

Preferably, the expression vector contains a selectable marker gene to allow the selection of transformed host cells. Selection genes are well known in the art and will vary with the host cell used. Expression and cloning vectors will typically contain a selection gene, also termed a selectable

marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, e.g., ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available for from complex media, e.g., the gene encoding D-alanine racemase for Bacilli.

Host cells transformed with a nucleotide sequence encoding a NPP may be cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein produced by a recombinant cell may be secreted, membrane-bound, or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides encoding the NPP can be designed with signal sequences which direct secretion of the NPP through a prokaryotic or eukaryotic cell membrane. The desired NPP may be produced recombinantly not only directly, but also as a fusion polypeptide with a heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the NPP-encoding DNA that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, e.g., the yeast invertase leader, alpha factor leader (including Saccharomyces and Kluyveromyces a-factor leaders, the latter described in U.S. Pat. No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published Apr. 4, 1990), or the signal described in WO 90113646 published Nov. 15, 1990. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders. According to the expression system selected, the coding sequence is inserted into an appropriate vector, which in turn may require the presence of certain characteristic "control elements" or "regulatory sequences." Appropriate constructs are known generally in the art (Ausubel, et al., 1990) and, in many cases, are available from commercial suppliers such as Invitrogen (San Diego, Calif.), Stratagene (La Jolla, Calif.), Gibco BRL (Rockville, Md.) or Clontech (Palo Alto, Calif.).

Expression in Bacterial Systems

Transformation of bacterial cells may be achieved using an inducible promoter such as the hybrid lacZ promoter of the "BLUESCRIPT" Phagemid (Stratagene) or "pSPORT1" (Gibco BRL). In addition, a number of expression vectors may be selected for use in bacterial cells to produce cleavable fusion proteins that can be easily detected and/or purified, including, but not limited to "BLUESCRIPT" (a-galactosidase; Stratagene) or pGEX (glutathione S-transferase; Promega,

Madison, Wis.). A suitable bacterial promoter is any nucleic acid sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of the coding sequence of the NPP into mRNA. A bacterial promoter has a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region typically includes an RNA polymerase binding site and a transcription initiation site. Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose and maltose, and sequences derived from biosynthetic enzymes such as tryptophan. Promoters from bacteriophage may also be used and are known in the art. In addition, synthetic promoters and hybrid promoters are also useful; for example, the tat promoter is a hybrid of the trp and lac promoter sequences. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. An efficient ribosome binding site is also desirable. The expression vector may also include a signal peptide sequence that provides for secretion of the NPP in bacteria. The signal sequence typically encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell, as is well known in the art. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). The bacterial expression vector may also include a selectable marker gene to allow for the selection of bacterial strains that have been transformed. Suitable selection genes include drug resistance genes such as ampicillin, chloramphenicol, erythromycin, kanamycin, neomycin and tetracycline. Selectable markers also include biosynthetic genes, such as those in the histidine, tryptophan and leucine biosynthetic pathways. When large quantities of a NPP are needed, e.g., for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be desirable. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the NPP coding sequence may be ligated into the vector in-frame with sequences for the amino-terminal Met and the subsequent 7 residues of beta-galactosidase so that a hybrid protein is produced; PIN vectors (Van Heeke & Schuster J Biol Chem 264:5503-5509 1989)); PET vectors (Novagen, Madison Wis.); and the like. Expression vectors for bacteria include the various components set forth above, and are well known in the art. Examples include vectors for *Bacillus subtilis, E. coli, Streptococcus cremoris,* and *Streptococcus lividans,* among others. Bacterial expression vectors are transformed into bacterial host cells using techniques well known in the art, such as calcium chloride mediated transfection, or electroporation.

Expression in Yeast

Yeast expression systems are well known in the art, and include expression vectors for *Saccharomyces cerevisiae, Candida albicans* and *C. maltosa, Hansenula polymorpha, Kluyveromyces fragilis* and *K. lactis, Pichia guillermondii* and *P pastoris, Schizosaccharomyces pombe,* and *Yarrowia lipolytica.* Examples of suitable promoters for use in yeast hosts include the promoters for 3-phosphoglycerate kinase (Hitzeman et al., *J. Biol. Chem.* 255:2073 (1980)) or other glycolytic enzymes (Hess et al., *J. Adv. Enzyme Reg.* 7:149 (1968); Holland, *Biochemistry* 17:4900 (1978)), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose- 6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, alpha factor, the ADH2IGAPDH promoter, glucokinase alcohol oxidase, and PGH. See, for example, Ausubel, et al., 1990; Grant et al., *Methods in Enzymology* 153:516-544, (1987). Other yeast promoters, which are inducible have the additional advantage of transcription controlled by growth conditions, include the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657. Yeast selectable markers include ADE2. HIS4. LEU2. TRPl. and ALG7, which confers resistance to tunicamycin; the neomycin phosphotransferase gene, which confers resistance to G418; and the CUP1 gene, which allows yeast to grow in the presence of copper ions. Yeast expression vectors can be constructed for intracellular production or secretion of NPP from the DNA encoding the NPP of interest. For example, a selected signal peptide and the appropriate constitutive or inducible promoter may be inserted into suitable restriction sites in the selected plasmid for direct intracellular expression of the NPP. For secretion of the NPP, DNA encoding the NPP can be cloned into the selected plasmid, together with DNA encoding the promoter, the yeast alpha-factor secretory signal/leader sequence, and linker sequences (as needed), for expression of the NPP. Yeast cells, can then be transformed with the expression plasmids described above, and cultured in an appropriate fermentation media. The protein produced by such transformed yeast can then be concentrated by precipitation with 10% trichloroacetic acid and analyzed following separation by SDS-PAGE and staining of the gels with Coomassie Blue stain. The recombinant NPP can subsequently be isolated and purified from the fermentation medium by techniques known to those of skill in the art.

Expression in Mammalian Systems

The NPPs may be expressed in mammalian cells. Mammalian expression systems are known in the art, and include retroviral vector mediated expression systems. Mammalian host cells may be

transformed with any of a number of different viral-based expression systems, such as adenovirus, where the coding region can be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a nonessential El or E3 region of the viral genome results in a viable virus capable of expression of the polypeptide of interest in infected host cells. A preferred expression vector system is a retroviral vector system such as is generally described in PCT/US97/01019 and PCT/US97/101048. Suitable mammalian expression vectors contain a mammalian promoter which is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence for NPP into mRNA. A promoter will have a transcription-initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, using a located 25-30 base pairs upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element (enhancer element), typically located within 100 to 200 base pairs upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation. Of particular use as mammalian promoters are the promoters from mammalian viral genes, since the viral genes are often highly expressed and have a broad host range. Examples include promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211, 504 published Jul. 5,1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, e.g., the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems. Transcription of DNA encoding a NPP by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA (usually about 10 to 300 bp) that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, a-fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer is preferably located at a site 5' from the promoter. In general, the transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-translational cleavage and polyadenylation. Examples of transcription terminator and polyadenylation signals include those derived from SV40. Long term, high-yield production of recombinant proteins can be effected in a stable expression system. Expression vectors which contain viral origins of replication or endogenous expression elements and a selectable marker gene may be

used for this purpose. Appropriate vectors containing selectable markers for use in mammalian cells are readily available commercially and are known to persons skilled in the art. Examples of such selectable markers include, but are not limited to herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase for use in tk- or hprt-cells, respectively. The methods of introducing exogenous nucleic acid into mammalian hosts, as well as other hosts, is well known in the art, and will vary with the host cell used. Techniques include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, viral infection, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

NPP can be purified from culture supernatants of mammalian cells transiently transfected or stably transformed by an expression vector carrying a NPP-encoding sequence. NPP is purified from culture supernatants of COS 7 cells transiently transfected by the pcD expression vector. Transfection of COS 7 cells with pcD proceeds as follows: One day prior to transfection, approximately 10^6 COS 7 monkey cells are seeded onto individual 100 mm plates in Dulbecco's modified Eagle medium (DME) containing 10% fetal calf serum and 2 mM glutamine. To perform the transfection, the medium is aspirated from each plate and replaced with 4 ml of DME containing 50 mM Tris.HCl pH 7.4, 400 mg/ml DEAE-Dextran and 50 μg of plasmid DNA. The plates are incubated for four hours at 37°C, then the DNA-containing medium is removed, and the plates are washed twice with 5 ml of serum-free DME. DME is added back to the plates which are then incubated for an additional 3 hrs at 37°C. The plates are washed once with DME, after which DME containing 4% fetal calf serum, 2 mM glutamine, penicillin (100 U/L) and streptomycin (100 μg/L) at standard concentrations is added. The cells are then incubated for 72 hrs at 37°C, after which the growth medium is collected for purification of NPP. Alternatively, transfection can be accomplished by electroporation. Plasmid DNA for the transfections is obtained by growing pcD(SRα), or like expression vector, containing the NPP-encoding cDNA insert in E. coli MC1061, described by Casadaban and Cohen, J. Mol. Biol., Vol. 138, pgs. 179-207 (1980), or like organism. The plasmid DNA is isolated from the cultures by standard techniques, e.g. Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor Laboratory, New York, 1989) or Ausubel et al (1990, cited above).

Expression in Insect Cells

NPPs may also be produced in insect cells. Expression vectors for the transformation of insect cells, and in particular, baculovirus-based expression vectors, are well known in the art. In one such system, the NPP-encoding DNA is fused upstream of an epitope tag contained within a baculovirus

expression vector. *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* Sf9 cells or in Trichoplusia larvae. The NPP-encoding sequence is cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of a NPP-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein coat. The recombinant viruses are then used to infect *S. frugiperda* cells or Trichoplusia larvae in which the NPP is expressed (Smith et al., J. Wol. 46:584 (1994); Engelhard E K et al., *Proc. Nat. Acad. Sci.* 91:3224-3227 (1994)). Suitable epitope tags for fusion to the NPP-encoding DNA include poly-his tags and immunoglobulin tags (like Fc regions of IgG). A variety of plasmids may be employed, including commercially available plasmids such as pVL1393 (Novagen). Briefly, the NPP-encoding DNA or the desired portion of the NPP-encoding DNA is amplified by PCR with primers complementary to the 5' and 3' regions. The 5' primer may incorporate flanking restriction sites. The PCR product is then digested with the selected restriction enzymes and subcloned into an expression vector. Recombinant baculovirus is generated by co-transfecting the above plasmid and BaculoGoldTM virus DNA (Pharmingen) into *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711) using lipofectin (commercially available from GIBCO-BRL), or other methods known to those of skill in the art. Virus is produced by day 4-5 of culture in Sf9 cells at 28°C, and used for further amplifications. Procedures are performed as further described in O'Reilley et al., *BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL,* Oxford University Press (1994). Extracts may be prepared from recombinant virus-infected Sf9 cells as described in Rupert et al., *Nature* 362:175-179 (1993). Alternatively, expressed epitope-tagged NPPs can be purified by affinity chromatography, or for example, purification of an IgG tagged (or Fc tagged) NPP can be performed using chromatography techniques, including Protein A or protein G column chromatography.

Evaluation of Gene Expression

Gene expression may be evaluated in a sample directly by standard techniques, e.g., Genechips® (Affymetrix, Santa Clara, CA) or Northern blotting (to determine the transcription of mRNA), dot blotting (DNA or RNA), or in situ hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be used in assays for detection of polypeptides, nucleic acids, such as specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. Such antibodies may be labeled and the assay carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected. Gene expression, alternatively, may be measured by immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to directly evaluate the expression of a NPP or encoding polynucleotide. Antibodies

useful for such immunological assays may be either monoclonal or polyclonal, and may be prepared against a native sequence NPP. Protein levels may also be detected by mass spectrometry. A further method of protein detection is with protein chips.

Purification of Expressed Protein

Expressed NPPs may be purified or isolated after expression, using any of a variety of methods known to those skilled in the art. The appropriate technique will vary depending upon what other components are present in the sample. Contaminant components that are removed by isolation or purification are materials that would typically interfere with diagnostic or therapeutic uses for the polypeptide, and may include enzymes, hormones, and other solutes. The purification step(s) selected will depend, for example, on the nature of the production process used and the particular NPP produced. As NPPs are secreted, they may be recovered from culture medium. Alternatively, the NPPs may be recovered from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (e.g. Triton-X 100) or by enzymatic cleavage. Alternatively, cells employed in expression of NPP can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or by use of cell lysing agents. Exemplary purification methods include, but are not limited to, ion-exchange column chromatography; chromatography using silica gel or a cation-exchange resin such as DEAE; gel filtration using, for example, Sephadex G-75; protein A Sepharose columns to remove contaminants such as IgG; chromatography using metal chelating columns to bind epitope-tagged forms of the NPP; ethanol precipitation; reverse phase HPLC; chromatofocusing; SDS-PAGE; and ammonium sulfate precipitation. Ordinarily, an isolated NPP will be prepared by at least one purification step. For example, the NPP may be purified using a standard anti-NPP antibody column. Ultrafiltration and dialysis techniques, in conjunction with protein concentration, are also useful (Scopes, R., *PROTEIN PURIFICATION,* Springer-Verlag, New York, N.Y., 1982). The degree of purification necessary will vary depending on the use of the NPP. In some instances no purification will be necessary. Once expressed and purified as needed, the NPP and encoding nucleic acids of the present invention are useful in a number of applications, as detailed herein.

Transgenic animals

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NPP-encoding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NPP sequences have been introduced into their genome or homologous recombinant animals in which endogenous NPP sequences have been

altered. Such animals are useful for studying the function and/or activity of NPPs or fragment thereof and for identifying and/or evaluating modulators of NPP biological activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal include a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a NPP-encoding nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection or retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The NPP cDNA sequence or a fragment thereof can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human NPP gene, such as from mouse or rat, can be used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a NPP transgene to direct expression of a NPP to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Pat. Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Pat. No. 4,873,191 by Wagner et al. and in Hogan, B., Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of a NPP transgene in its genome and/or expression of NPP mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a NPP can further be bred to other transgenic animals carrying other transgenes.

To create an animal in which a desired nucleic acid has been introduced into the genome via homologous recombination, a vector is prepared which contains at least a portion of a NPP gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the NPP gene. The NPP gene can be a human gene, but more preferably, is a non-human homologue (e.g., a cDNA isolated by stringent hybridization with a nucleotide sequence coding for a NPP). For example, a mouse NPP gene can be used to construct a homologous recombination vector

suitable for altering an endogenous gene in the mouse genome. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous NPP gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NPP gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous NPP). In the homologous recombination vector, the altered portion of the NPP gene is flanked at its 5' and 3' ends by additional nucleic acid sequence of the NPP gene to allow for homologous recombination to occur between the exogenous sequence carried by the vector and an endogenous gene in an embryonic stem cell. The additional flanking nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K. R. and Capecchi, M. R. (1987) Cell 51:503, for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced NPP gene has homologously recombined with the endogenous gene are selected (see e.g., Li, E. et al. (1992) Cell 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in Teratocarcinomas and Embryonic Stem Cells. A Practical Approach, E. J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) Current Opinion in Biotechnology 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec et al.; WO 91/01140 by Smithies et al.; WO 92/0968 by Zijlstra et al.; and WO 93/04169 by Berns et al.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, see, e.g., Lakso et al. (1992) PNAS 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of Saccharomyces cerevisiae (O'Gorman et al. (1991) Science 251:1351-1355). If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected

protein and the other containing a transgene encoding a recombinase.

Assessing NPP activity

It will be appreciated that the invention further provides methods of testing the activity of, or obtaining, functional fragments and variants of NPPs. Such methods involve providing a variant or modified NPP or NPP-encoding nucleic acid and assessing whether the polypeptide displays a NPP biological activity. Encompassed is thus a method of assessing the function of a NPP comprising: (a) providing a NPP, or a biologically active fragment or homologue thereof; and (b) testing said NPP, or a biologically active fragment or homologue thereof for a NPP biological activity. Any suitable format may be used, including cell free, cell-based and in vivo formats. For example, said assay may comprise expressing a NPP nucleic acid in a host cell, and observing activity in said cell and other affected cells.

A NPP biological activity may be any activity as described herein, such as (1) indicating that an individual has or will have a cardiovascular disorder; (2) circulating through the bloodstream of individuals with a cardiovascular disorder; (3) antigenicity, or the ability to bind an anti-NPP specific antibody; (4) immunogenicity, or the ability to generate an anti-NPP specific antibody; (5) forming intermolecular amino acid side chain interactions such as hydrogen, amide, or preferably disulfide links; (6) interaction with a NPP target molecule; and (7) undergoing posttranslational processing, for example, specific protealysis.

NPP biological activity can be assayed by any suitable method known in the art. Antigenicity and immunogenicity may be detected, for example, as described in the sections titled "Anti NPP antibodies" and "Uses of NPP antibodies." Circulation in blood plasma may be detected as described in "Diagnostic and Prognostic Uses."

Determining the ability of the NPP to bind to or interact with a NPP target molecule can be accomplished by a method for directly or indirectly determining binding, as is common to the art. Such methods can be cell-based or cell free. Interaction of a test compound with a NPP can be detected, for example, by coupling the NPP or biologically active portion thereof with a label group such that binding of the NPP or biologically active portion thereof to its cognate target molecule can be determined by detecting the labeled NPP or biologically active portion thereof in a complex. For example, the extent of complex formation may be measured by immunoprecipitating the complex or by performing gel electrophoresis. Determining the ability of the NPP to bind to a NPP target molecule may also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) Anal. Chem. 63:2338-2345 and Szabo et al. (1995) Curr. Opin. Struct. Biol. 5:699-705, the disclosures of which are incorporated herein by reference in their entireties. As used herein, "BIA" is a technology for studying biospecific

interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules. Protein array methods are useful for detecting interaction (e.g., Proteinchip®, Ciphergen Biosystems, Fremont, CA). For example, one member of a receptor/ ligand pair is docked to an adsorbent, and its ability to bind the binding partner is determined in the presence of the test substance. Because of the rapidity with which adsorption can be tested, combinatorial libraries of test substances can be easily screened for their ability to modulate the interaction. In preferred methods, NPPs are docked to the adsorbent. Binding partners are preferably labeled, thus enabling detection of the interaction. Alternatively, in certain embodiments, a test substance is docked to the adsorbent. The polypeptides of the invention are exposed to the test substance and binding detected.

Cardiovascular disorders may be diagnosed by any method determined appropriate for an individual by one of skill in the art. Further examples of symptoms and diagnostics may be found in the Background section, and are best determined appropriately by one of skill in the art based on the particular profile of a patient.

Specific proteolysis may be detected by comparing the molecular weight of a sample peptide to that of a peptide of known molecular weight. Molecular weights are easily compared according to any method common to the art such as SDS-PAGE, gel chromatography, or mass spectrometry. Preferably, the molecular weight of a test peptide is obtained by mass spectrometry.

Anti-NPP antibodies

The present invention provides antibodies and binding compositions specific for NPPs. Such antibodies and binding compositions include polyclonal antibodies, monoclonal antibodies, Fab and single chain Fv fragments thereof, bispecific antibodies, heteroconjugates, and humanized antibodies. Such antibodies and binding compositions may be produced in a variety of ways, including hybridoma cultures, recombinant expression in bacteria or mammalian cell cultures, recombinant expression in transgenic animals, and the like. There is abundant guidance in the literature for selecting a particular production methodology, e.g. Chadd and Chamow, Curr. Opin. Biotechnol., 12: 188-194 (2001).

The choice of manufacturing methodology depends on several factors including the antibody structure desired, the importance of carbohydrate moieties on the antibodies, ease of culturing and purification, cost, and the like. Many different antibody structures may be generated using standard expression technology, including full-length antibodies, antibody fragments, such as Fab and Fv fragments, as well as chimeric antibodies comprising components from different species. Antibody fragments of small size, such as Fab and Fv fragments, having no effector functions and limited pharmokinetic activity may be generated in a bacterial expression system. Single chain Fv fragments

are highly selective for in vivo tumors, show good tumor penetration and low immunogenicity, and are cleared rapidly from the blood, e.g. Freyre et al, J. Biotechnol., 76: 157-163 (2000). Thus, such molecules are desirable for radioimmunodetection and in situ radiotherapy. Whenever pharmacokinetic activity in the form of increased half-life is required for therapeutic purposes, then full-length antibodies are preferable. For example, immunoglobulin G (IgG) the molecule may be one of four subclasses: γ1, γ2, γ3, or γ4. If a full-length antibody with effector function is required, then IgG subclasses γ1 or γ3 are preferred, and IgG subclass γ1 is most preferred. The γ1 and γ3 subclasses exhibit potent effector function, complement activation, and promote antibody-dependent cell-mediated cytotoxicity through interaction with specific Fc receptors, e.g. Raju et al, Glycobiology, 10: 477-486 (2000); Lund et al, J. Immunol., 147: 2657-2662 (1991).

Polyclonal Antibodies

The anti-NPP antibodies of the present invention may be polyclonal antibodies. Such polyclonal antibodies can be produced in a mammal, for example, following one or more injections of an immunizing agent, and preferably, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected into the mammal by a series of subcutaneous or intraperitoneal injections. The immunizing agent may include a NPP or a fusion protein thereof. It may be useful to conjugate the antigen to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include, but are not limited to, keyhole limpet hemocyanin (KLH), methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Adjuvants include, for example, Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicoryno-mycolate). The immunization protocol may be determined by one skilled in the art based on standard protocols or by routine experimentation.

Alternatively, a crude protein preparation which has been enriched for a NPP or a portion thereof can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies are purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate and excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appear to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987). An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. J.

Clin. Endocrinol. Metab. 33:988-991(1971). Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum. Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D. C. (1980).

Monoclonal Antibodies

Alternatively, the anti-NPP antibodies may be monoclonal antibodies. Monoclonal antibodies may be produced by hybridomas, wherein a mouse, hamster, or other appropriate host animal, is immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent, e.g. Kohler and Milstein, Nature 256:495 (1975). The immunizing agent will typically include the NPP or a fusion protein thereof and optionally a carrier. Alternatively, the lymphocytes may be immunized *in vitro*. Generally, spleen cells or lymph node cells are used if non-human mammalian sources are desired, or peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired. The lymphocytes are fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to produce a hybridoma cell, e.g. Goding, *MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE,* Academic Press, pp. 59-103 (1986); Liddell and Cryer, A Practical Guide to Monoclonal Antibodies (John Wiley & Sons, New York, 1991); Malik and Lillenoj, Editors, Antibody Techniques (Academic Press, New York, 1994). In general, immortalized cell lines are transformed mammalian cells, for example, myeloma cells of rat, mouse, bovine or human origin. The hybridoma cells are cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine (HAT), substances which prevent the growth of HGPRT-deficient cells. Preferred immortalized cell lines are those that fuse efficiently, support stable high level production of antibody, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine or human myeloma lines, which can be obtained, for example, from the American Type Culture Collection (ATCC), Rockville, MD. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies, e.g. Kozbor, J. Immunol. 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, pp. 51-63 (1987).

The culture medium (supernatant) in which the hybridoma cells are cultured can be assayed for the presence of monoclonal antibodies directed against a NPP. Preferably, the binding specificity of monoclonal antibodies present in the hybridoma supernatant is determined by immunoprecipitation or by an in vitro binding assay, such as radio- immunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Appropriate techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, *Anal. Biochem.* 107:220 (1980). After the desired antibody-producing hybridoma cells are identified, the cells may be cloned by limiting dilution procedures and grown by standard methods (Goding, 1986, *supra*). Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown in vivo as ascites in a mammal. The monoclonal antibodies secreted by selected clones may be isolated or purified from the culture medium or ascites fluid by immunoglobulin purification procedures routinely used by those of skill in the art such as, for example, protein A-Sepharose, hydroxyl-apatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Pat. No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be isolated from the NPP-specific hybridoma cells and sequenced, e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies. Once isolated, the DNA may be inserted into an expression vector, which is then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for the murine heavy and light chain constant domains for the homologous human sequences (Morrison et al., *Proc. Nat. Acad. Sci.* 81:6851-6855 (1984); Neuberger et al., *Nature* 312:604-608 (1984); Takeda et al., *Nature* 314:452-454 (1985)), or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. The non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody. The antibodies may also be monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, in vitro methods are suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art.

Antibodies and antibody fragments characteristic of hybridomas of the invention can also be

produced by recombinant means by extracting messenger RNA, constructing a cDNA library, and selecting clones which encode segments of the antibody molecule. The following are exemplary references disclosing recombinant techniques for producing antibodies: Wall et al., Nucleic Acids Research, Vol. 5, pgs. 3113-3128 (1978); Zakut et al., Nucleic Acids Research, Vol. 8, pgs. 3591-3601 (1980); Cabilly et al., Proc. Natl. Acad. Sci., Vol. 81, pgs. 3273-3277 (1984); Boss et al., Nucleic Acids Research, Vol. 12, pgs. 3791-3806 (1984); Amster et al., Nucleic Acids Research, Vol. 8, pgs. 2055-2065 (1980); Moore et al., U.S. Patent 4,642,334; Skerra et al, Science, Vol. 240, pgs. 1038-1041(1988); Huse et al, Science, Vol. 246, pgs. 1275-1281 (1989); and U.S. patents 6,054,297; 5,530,101; 4,816,567; 5,750,105; and 5,648,237. In particular, such techniques can be used to produce interspecific monoclonal antibodies, wherein the binding region of one species is combined with non-binding region of the antibody of another species to reduce immunogenicity, e.g. Liu et al., Proc. Natl. Acad. Sci., Vol. 84, pgs. 3439-3443 (1987), and US patents 6,054,297 and 5,530,101. Preferably, recombinantly produced Fab and Fv fragments are expressed in bacterial host systems. Preferably, full-length antibodies are produced by mammalian cell culture techniques. More preferably, full-length antibodies are expressed in Chinese Hamster Ovary (CHO) cells or NSO cells.

Both polyclonal and monoclonal antibodies can be screened by ELISA. As in other solid phase immunoassays, the test is based on the tendency of macromolecules to adsorb nonspecifically to plastic. The irreversibility of this reaction, without loss of immunological activity, allows the formation of antigen-antibody complexes with a simple separation of such complexes from unbound material. To titrate antipeptide serum, peptide conjugated to a carrier different from that used in immunization is adsorbed to the wells of a 96-well microtiter plate. The adsorbed antigen is then allowed to react in the wells with dilutions of anti-peptide serum. Unbound antibody is washed away, and the remaining antigen-antibody complexes are allowed to react with an antibody specific for the IgG of the immunized animal. This second antibody is conjugated to an enzyme such as alkaline phosphatase. A visible colored reaction produced when the enzyme substrate is added indicates which wells have bound antipeptide antibodies. The use of spectrophotometer readings allows better quantification of the amount of peptide-specific antibody bound. High-titer antisera yield a linear titration curve between 10^{-3} and 10^{-5} dilutions.

NPP carriers

The invention includes immunogens derived from NPPs and immunogens comprising conjugates between carriers and peptides of the invention. The term immunogen as used herein refers to a substance which is capable of causing an immune response. The term carrier as used herein refers to any substance which when chemically conjugated to a peptide of the invention permits a host organism immunized with the resulting conjugate to generate antibodies specific for the conjugated

peptide. Carriers include red blood cells, bacteriophages, proteins, or synthetic particles such as agarose beads. Preferably, carriers are proteins, such as serum albumin, gamma-globulin, keyhole limpet hemocyanin, thyroglobulin, ovalbumin, fibrinogen, or the like.

The general technique of linking synthetic peptides to a carrier is described in several references, e.g. Walter and Doolittle, "Antibodies Against Synthetic Peptides," in Setlow et al., eds., Genetic Engineering, Vol. 5, pgs. 61-91 (Plenum Press, N.Y., 1983); Green et al. Cell, Vol. 28, pgs. 477-487 (1982); Lerner et al., Proc. Natl. Acad. Sci., 78:3403-3407 (1981); Shimizu et al., U.S. Patent 4,474,754; and Ganfield et al., U.S. Patent 4,311,639. Also, techniques employed to link haptens to carriers are essentially the same as the above-referenced techniques, e.g. chapter 20 in Tijssen, Practice and Theory of Enzyme Immunoassays (Elsevier, New York, 1985). The four most commonly used schemes for attaching a peptide to a carrier are (1) glutaraldehyde for amino coupling, e.g. as disclosed by Kagan and Glick, in Jaffe and Behrman, eds. Methods of Hormone Radioimmunoassay, pgs. 328-329 (Academic Press, N.Y., 1979), and Walter et al. Proc. Natl. Acad. Sci., 77:5197-5200 (1980); (2) water-soluble carbodiimides for carboxyl to amino coupling, e.g. as disclosed by Hoare et al., J. Biol. Chem., 242:2447-2453 (1967); (3) bis-diazobenzidine (BDB) for tyrosine to tyrosine sidechain coupling, e.g. as disclosed by Bassiri et al., pgs. 46-47, in Jaffe and Behrman, eds. (cited above), and Walter et al. (cited above); and (4) maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) for coupling cysteine (or other sulfhydryls) to amino groups, e.g. as disclosed by Kitagawa et al., J. Biochem. (Tokyo), 79:233-239 (1976), and Lerner et al. (cited above). A general rule for selecting an appropriate method for coupling a given peptide to a protein carrier can be stated as follows: the group involved in attachment should occur only once in the sequence, preferably at the appropriate end of the segment. For example, BDB should not be used if a tyrosine residue occurs in the main part of a sequence chosen for its potentially antigenic character. Similarly, centrally located lysines rule out the glutaraldehyde method, and the occurrences of aspartic and glutamic acids frequently exclude the carbodiimide approach. On the other hand, suitable residues can be positioned at either end of chosen sequence segment as attachment sites, whether or not they occur in the "native" protein sequence. Internal segments, unlike the amino and carboxy termini, will differ significantly at the "unattached end" from the same sequence as it is found in the native protein where the polypeptide backbone is continuous. The problem can be remedied, to a degree, by acetylating the α-amino group and then attaching the peptide by way of its carboxy terminus. The coupling efficiency to the carrier protein is conveniently measured by using a radioactively labeled peptide, prepared either by using a radioactive amino acid for one step of the synthesis or by labeling the completed peptide by the iodination of a tyrosine residue. The presence of tyrosine in the peptide also allows one to set up a sensitive radioimmune assay, if desirable. Therefore, tyrosine can be introduced as a terminal residue if it is not part of the peptide sequence defined by the native

polypeptide.

Preferred carriers are proteins, and preferred protein carriers include bovine serum albumin, myoglobulin, ovalbumin (OVA), or keyhole limpet hemocyanin (KLH). Peptides can be linked to KLH through cysteines by MBS as disclosed by Liu et al., Biochemistry, Vol. 18, pgs. 690-697 (1979). The peptides are dissolved in phosphate-buffered saline (pH 7.5), 0.1 M sodium borate buffer (pH 9.0) or 1.0 M sodium acetate buffer (pH 4.0). The pH for the dissolution of the peptide is chosen to optimize peptide solubility. The content of free cysteine for soluble peptides is determined by Ellman's method, Ellman, Arch. Biochem. Biophys., Vol. 82, pg. 7077 (1959). For each peptide, 4 mg KLH in 0.25 ml of 10 mM sodium phosphate buffer (pH 7.2) is reacted with 0.7 mg MBS (dissolved in dimethyl formamide) and stirred for 30 min at room temperature. The MBS is added dropwise to ensure that the local concentration of formamide is not too high, as KLH is insoluble in >30% formamide. The reaction product, KLH-MBS, is then passed through Sephadex G-25 equilibrated with 50 mM sodium phosphate buffer (pH 6.0) to remove free MBS, KLH recovery from peak fractions of the column eluate (monitored by OD280) is estimated to be approximately 80%. KLH-MBS is then reacted with 5 mg peptide dissolved in 1 ml of the chosen buffer. The pH is adjusted to 7-7.5 and the reaction is stirred for 3 hr at room temperature. Coupling efficiency is monitored with radioactive peptide by dialysis of a sample of the conjugate against phosphate-buffered saline, and may range from 8% to 60%. Once the peptide-carrier conjugate is available polyclonal or monoclonal antibodies are produced by standard techniques, e.g. as disclosed by Campbell, Monoclonal Antibody Technology (Elsevier, New York, 1984); Hurrell, ed. Monoclonal Hybridoma Antibodies: Techniques and Applications (CRC Press, Boca Raton, FL, 1982); Schreier et al. Hybridoma Techniques (Cold Spring Harbor Laboratory, New York, 1980); or U.S. Patent 4,562,003.

Humanized Antibodies

The anti-NPP antibodies of the invention may further comprise humanized antibodies or human antibodies. The term "humanized antibody" refers to humanized forms of non-human (e.g., murine) antibodies that are chimeric antibodies, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab'), or other antigen-binding partial sequences of antibodies) which contain some portion of the sequence derived from non-human antibody. Humanized antibodies include human immunoglobulins in which residues from a complementary determining region (CDR) of the human immunoglobulin are replaced by residues from a CDR of a non-human species such as mouse, rat or rabbit having the desired binding specificity, affinity and capacity. In general, the humanized antibody will comprise substantially all of at least one, and generally two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or

substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., Nature 321:522-525 (1986) and Presta, Curr. Op. Struct. Biol. 2:593-596 (1992)). Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acids introduced into it from a source which is non-human in order to more closely resemble a human antibody, while still retaining the original binding activity of the antibody. Methods for humanization of antibodies are further detailed in Jones et al., *Nature* 321:522-525 (1986); Riechmann et al., *Nature* 332:323-327 (1988); and Verhoeyen et al., *Science* 239:1534-1536 (1988). Such "humanized" antibodies are chimeric antibodies in that substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species.

Heteroconjugate Antibodies

Heteroconjugate antibodies which comprise two covalently joined antibodies, are also within the scope of the present invention. Heteroconjugate antibodies may be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins may be prepared using a disulfide exchange reaction or by forming a thioether bond.

Bispecific Antibodies

Bispecific antibodies have binding specificities for at least two different antigens. Such antibodies are monoclonal, and preferably human or humanized. One of the binding specificities of a bispecific antibody of the present invention is for a NPP, and the other one is preferably for a cell-surface protein or receptor or receptor subunit. Methods for making bispecific antibodies are known in the art, and in general, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs in hybridoma cells, where the two heavy chains have different specificities, e.g. Milstein and Cuello, *Nature 305:537-539* (1983). Given that the random assortment of immunoglobulin heavy and light chains results in production of potentially ten different antibody molecules by the hybridomas, purification of the correct molecule usually requires some sort of affinity purification, e.g. affinity chromatography.

Uses of NPP Antibodies

Anti-NPP antibodies are preferably specific for the NPPs of the invention and as such, do not bind peptides derived from other proteins with high affinity. As used herein, the term "heavy chain variable region" means a polypeptide (1) which is from 110 to 125 amino acids in length, and (2) whose amino acid sequence corresponds to that of a heavy chain of an antibody of the invention,

starting from the heavy chain's N-terminal amino acid. Likewise, the term "light chain variable region" means a polypeptide (1) which is from 95 to 115 amino acids in length, and (2) whose amino acid sequence corresponds to that of a light chain of an antibody of the invention, starting from the light chain's N-terminal amino acid. As used herein the term "monoclonal antibody" refers to homogeneous populations of immunoglobulins which are capable of specifically binding to NPP.

The use of antibody fragments is also well known, e.g. Fab fragments: Tijssen, Practice and Theory of Enzyme Immunoassays (Elsevier, Amsterdam, 1985); and Fv fragments: Hochman et al. Biochemistry, 12:1130-1135 (1973), Sharon et al., Biochemistry, 15:1591-1594 (1976) and Ehrlich et al., U.S. Patent 4,355,023; and antibody half molecules: Auditore- Hargreaves, U.S. Patent 4,470,925.

Preferably, monoclonal antibodies, Fv fragments, Fab fragments, or other binding compositions derived from monoclonal antibodies of the invention have a high affinity to NPPs. The affinity of monoclonal antibodies and related molecules to NPP may be measured by conventional techniques including plasmon resonance, ELISA, and equilibrium dialysis. Affinity measurement by plasmon resonance techniques may be carried out, for example, using a BIAcore 2000 instrument (Biacore AB, Uppsala, Sweden) in accordance with the manufacturer's recommended protocol. Preferably, affinity is measured by ELISA, for example, as described in U.S. patent 6,235,883. Preferably, the dissociation constant between NPPs and monoclonal antibodies of the invention is less than 10^{-5} molar. More preferably, such dissociation constant is less than 10^{-8} molar; still more preferably, such dissociation constant is less than 10^{-9} molar; and most preferably, such dissociation constant is in the range of 10^{-9} to 10^{-11} molar.

The antibodies of the present invention are useful for detection. Such detection methods are advantageously applied to diagnosis and prognosis. The antibodies of the invention may be used in most assays involving antigen-antibody reactions. The assays may be homogeneous or heterogeneous. In a homogeneous assay approach, the sample can be a biological sample or fluid such as serum, urine, whole blood, lymphatic fluid, plasma, saliva, cells, tissue, and material secreted by cells or tissues cultured in vitro. The sample can be pretreated if necessary to remove unwanted materials. The immunological reaction usually involves the antigen-specific antibody, labeled antigen, and the sample suspected of containing the antigen. The antigen can be directly labeled with any label group described herein.

In a heterogeneous assay approach, the reagents are usually the sample, the specific antibody, and means for producing a detectable signal. The specimen is generally placed on a support, such as a plate or a slide, and contacted with the antibody in a liquid phase. The support is then separated from the liquid phase and either the support phase or the liquid phase is examined for a detectable signal employing means for producing such signal or signal producing system. The signal is related to the

presence of the antigen in the sample. Means for producing a detectable signal includes the use of any label group described herein.

One embodiment of an assay employing an antibody of the present invention involves the use of a surface to which the monoclonal antibody of the invention is attached. The underlying structure of the surface may take different forms, have different compositions and may be a mixture of compositions or laminates or combinations thereof. The surface may assume a variety of shapes and forms and may have varied dimensions, depending on the manner of use and measurement. Illustrative surfaces may be pads, beads, discs, or strips which may be flat, concave or convex. Thickness is not critical, generally being from about 0.1 to 2 mm thick and of any convenient diameter or other dimensions. The surface typically will be supported on a rod, tube, capillary, fiber, strip, disc, plate, cuvette and will typically be porous and polyfunctional or capable of being polyfunctionalized so as to permit covalent binding of an antibody and permit bonding of other compounds which form a part of a means for producing a detectable signal. A wide variety of organic and inorganic polymers, both natural and synthetic, and combinations thereof, may be employed as the material for the solid surface. Illustrative polymers include polyethylene, polypropylene, poly(4-methylbutene), polystyrene, polymethracrylate, poly(ethylene terephthalate), rayon, nylon, poly(vinyl butyrate), silicones, polyformaldehyde, cellulose, cellulose acetate, nitrocellulose, and latex. Other surfaces include paper, glasses, ceramics, metals, metaloids, semiconductor materials, cements, silicates or the like. Also included are substrates that form gels, gelatins, lipopolysaccharides, silicates, agarose and polyacrylamides or polymers which form several aqueous phases such as dextrans, polyalkylene glycols (alkylene of 2 to 3 carbon atoms) or surfactants such as phospholipids. The binding of the antibody to the surface may be accomplished by well known techniques, commonly available in the literature (See, for example, "Immobilized Enzymes," Ichiro Chibata, Press, New York (1978) and Cuatrecasas, J. Bio. Chem., 245: 3059 (1970)). In carrying out the assay in accordance with this aspect of the invention the sample is mixed with aqueous medium and the medium is contacted with the surface having an antibody bound thereto. Labels may be included in the aqueous medium, either concurrently or added subsequently so as to provide a detectable signal associated with the surface. The means for producing the detectable signal can involve the incorporation of a labeled analyte or it may involve the use of a second monoclonal antibody having a label conjugated thereto. Separation and washing steps will be carried out as needed. The signal detected is related to the presence of NPP in the sample. It is within the scope of the present invention to include a calibration on the same support. A particular embodiment of an assay in accordance with the present invention, by way of illustration and not limitation, involves the use of a support such as a slide or a well of a petri dish. The technique involves fixing the sample to be analyzed on the support with an appropriate fixing material and incubating the sample on the slide with a monoclonal antibody. After washing with an

appropriate buffer such as, for example, phosphate buffered saline, the support is contacted with a labeled specific binding partner for the antibody. After incubation as desired, the slide is washed a second time with an aqueous buffer and the determination is made of the binding of the labeled monoclonal antibody to the antigen. If the label is fluorescent, the slide may be covered with a fluorescent antibody mounting fluid on a cover slip and then examined with a fluorescent microscope to determine the extent of binding. On the other hand, the label can be an enzyme conjugated to the monoclonal antibody and the extent of binding can be determined by examining the slide for the presence of enzyme activity, which may be indicated by the formation of a precipitate, color, etc. A particular example of an assay utilizing the present antibodies is a double determinant ELISA assay. A support such as, e.g., a glass or vinyl plate, is coated with anti-NPP antibodies by conventional techniques. The support is contacted with the sample suspected of containing NPP, usually in aqueous medium. After an incubation period from 30 seconds to 12 hours, the support is separated from the medium, washed to remove unbound NPP with, for example, water or an aqueous buffered medium, and contacted with an antibody specific for NPP, again usually in aqueous medium. The antibody is labeled with an enzyme directly or indirectly such as, e.g., horseradish peroxidase or alkaline phosphatase. After incubation, the support is separated from the medium, and washed as above. The enzyme activity of the support or the aqueous medium is determined. This enzyme activity is related to the amount of NPP in the sample.

The invention also includes kits, e.g., diagnostic assay kits, for carrying out the methods disclosed above. In one embodiment, the kit comprises in packaged combination (a) a monoclonal antibody more specifically defined above and (b) a conjugate of a specific binding partner for the above monoclonal antibody and a label capable of producing a detectable signal. The reagents may also include ancillary agents such as buffering agents and protein stabilizing agents, e.g., polysaccharides and the like. The kit may further include, where necessary, other members of the signal producing system of which system the label is a member, agents for reducing background interference in a test, control reagents, apparatus for conducting a test, and the like. In another embodiment, the diagnostic kit comprises a conjugate of monoclonal antibody of the invention and a label capable of producing a detectable signal. Ancillary agents as mentioned above may also be present.

Further, an anti-NPP antibody can be used to isolate a NPP by standard techniques, such as affinity chromatography or immunoprecipitation. For example, an anti-NPP antibody can facilitate the purification of natural NPPs from cells and of recombinantly produced NPPs expressed in host cells. Moreover, an anti-NPP antibody can be used to isolate NPPs to aid in detection of low concentrations of NPP (e.g., in plasma, cellular lysate or cell supernatant) or in order to evaluate the abundance and pattern of expression. Anti-NPP antibodies can be used diagnostically to monitor protein levels in

tissue as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a label group.

Protein Arrays

Detection, purification, and screening of the polypeptides of the invention may be accomplished using retentate chromatography (preferably, protein arrays or chips), as described by U.S. Patent 6225027 and U.S. Patent Application 20010014461. Briefly, retentate chromatography describes methods in which polypeptides (and/ or other sample components) are retained on an adsorbent (e.g., array or chip) and subsequently detected. Such methods involve (1) selectively adsorbing polypeptides from a sample to a substrate under a plurality of different adsorbent/eluant combinations ("selectivity conditions") and (2) detecting the retention of adsorbed polypeptides by desorption spectrometry (e.g., by mass spectrometry). In conventional chromatographic methods, polypeptides are eluted off of the adsorbent prior to detection. The coupling of adsorption chromatography with detection by desorption spectrometry provides extraordinary sensitivity, the ability to rapidly analyze retained components with a variety of different selectivity conditions, and parallel processing of components adsorbed to different sites (i.e., "affinity sites" or "spots") on the array under different elution conditions.

These methods are useful for: combinatorial, biochemical separation and purification of the NPP; study of differential gene expression; detection of differences in protein levels (e.g., for diagnosis); and detection of molecular recognition events, (e.g., for screening and drug discovery). Thus, this invention provides a molecular discovery and diagnostic device that is characterized by the inclusion of both parallel and multiplex polypeptide processing capabilities. Polypeptides of the invention and NPP-binding substances are preferably attached to a label group, and thus directly detected, enabling simultaneous transmission of two or more signals from the same "circuit" (i.e., addressable "chip" location) during a single unit operation.

Detection of NPP by mass spectrometry

In accordance with the present invention, any instrument, method, process, etc. can be utilized to determine the identity and abundance of proteins in a sample. A preferred method of obtaining identity is by mass spectrometry, where protein molecules in a sample are ionized and then the resultant mass and charge of the protein ions are detected and determined.

To use mass spectrometry to analyze proteins, it is preferred that the protein be converted to a gas-ion phase. Various methods of protein ionization are useful, including, e.g., fast ion bombardment (FAB), plasma desorption, laser desorption, thermal desorption, preferably, electrospray ionization

(ESI) and matrix-assisted laser desorption/ionization (MALDI). Many different mass analyzers are available for peptide and protein analysis, including, but not limited to, Time-of-Flight (TOF), ion trap (ITMS), Fourier transform ion cyclotron (FTMS), quadrupole ion trap, and sector (electric and/or magnetic) spectrometers. See, e.g., U.S. Pat. No. 5,572,025 for an ion-trap MS. Mass analyzers can be used alone, or in combination with other mass analyzers in tandem mass spectrometers. In the latter case, a first mass analyzer can be use to separate the protein ions (precursor ion) from each other and determine the molecular weights of the various protein constituents in the sample. A second mass analyzer can be used to analyze each separated constituents, e.g., by fragmenting the precursor ions into product ions by using, e.g. an inert gas. Any desired combination of mass analyzers can be used, including, e.g., triple quadrupoles, tandem time-of-flights, ion traps, and/or combinations thereof.

Different kinds of detectors can be used to detect the protein ions. For example, destructive detectors can be utilized, such as ion electron multipliers or cryogenic detectors (e.g., U.S. Patent 5,640,010). Additionally, non-destructive detectors can be used, such as ion traps which are used as ion current pick-up devices in quadrupole ion trap mass analyzers or FTMS.

For MALDI-TOF, a number of sample preparation methods can be utilized including, dried droplet (Karasand Hillenkamp, Anal. Chem., 60:2299-2301, 1988), vacuum-drying (Winberger et al., In Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, May 31-June 4, 1993, pp. 775a-b), crush crystals (Xiang et al., Rapid Comm. Mass Spectrom., 8:199-204,1994), slow crystal growing (Xiang et al., Org. Mass Spectrom, 28:1424-1429, 1993); active film (Mock et al., Rapid Comm. Mass Spectrom.,6:233-238, 1992; Bai et al., Anal. Chem., 66:3423-3430, 1994), pneumatic spray (Kochling et al., Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics; Atlanta, GA, May 21-26, 1995, p1225); electrospray (Hensel et al., Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics; Atlanta, GA, May 21 -26, 1995, p947); fast solvent evaporation (Vorm et al., Anal. Chem., 66:3281-3287, 1994); sandwich (Li et al., J. Am. Chem. Soc., 11 8:11662-11663,1996); and two-layer methods (Dal et al., Anal. Chem., 71:1087-1091, 1999). See also, e.g., Liang et al., Rapid Commun. Mass Spectrom., 10: 1219-1226, 1996; van Adrichemet al., Anal. Chem., 70:923-930, 1998.

For MALDI analysis, samples are prepared as solid-state co-crystals or thin films by mixing them with an energy absorbing compound or colloid (the matrix) in the liquid phase, and ultimately drying the solution to the solid state upon the surface of an inert probe. In some cases an energy absorbing molecule (EAM) is an integral component of the sample presenting surface. Regardless of EAM application strategy, the probe contents are allowed to dry to the solid state prior to introduction into the laser desorption/ionization time-of-flight mass spectrometer (LDIMS).

Ion detection in TOF mass spectrometry is typically achieved with the use of electro-emissive

detectors such as electron multipliers (EMP) or microchannel plates (MCP). Both of these devices function by converting primary incident charged particles into a cascade of secondary, tertiary, quaternary, etc. electrons. The probability of secondary electrons being generated by the impact of a single incident charged particle can be taken to be the ion-to-electron conversion efficiency of this charged particle (or more simply, the conversion efficiency). The total electron yield for cascading events when compared to the total number of incident charged particles is typically described as the detector gain. Because generally the overall response time of MCPs is far superior to that of EMPs, MCPs are the preferred electro-emissive detector for enhancing mass/charge resolving power. However, EMPs function well for detecting ion populations of disbursed kinetic energies, where rapid response time and broad frequency bandwidth are not necessary.

In a preferred aspect, for the analysis of digested proteins, a liquid-chromatography tandem mass spectrometer (LC-TMS) is used. This system provides an additional stage of sample separation via use of a liquid chromatograph followed by tandem mass spectrometry.

The methods described herein of separating and fractionating proteins provide individual proteins or fractions containing small numbers of distinct proteins. These proteins can be identified by mass spectral determination of the molecular masses of the protein and peptides resulting from the fragmentation thereof. Making use of available information in protein sequence databases, a comparison can be made between proteolytic peptide mass patterns generated *in silico,* and experimentally observed peptide masses. Alternatively, a protein database can be constructed by carrying out a 6-frame translation of a nucleotide sequence database (e.g., Genbank). A "hit-list" can be compiled, ranking candidate proteins in the database, based on (among other criteria) the number of matches between the theoretical and experimental proteolytic fragments. Several Web sites are accessible that provide software for protein identification on-line, based on peptide mapping and sequence database search strategies (e.g., http://www.expasy.ch). Methods of peptide mapping and sequencing using MS are described in WO 95/252819, U.S. Pat. No. 5,538,897, U.S. Pat. No. 5,869,240, U.S. Pat. No. 5,572,259, and U.S. Pat. No. 5,696,376. See, also, Yates, J. Mass Spec., 33:1 (1998).

Data collected from a mass spectrometer typically comprises the intensity and mass to charge ratio for each detected event. Spectral data can be recorded in any suitable form, including, e.g., in graphical, numerical, or electronic formats, either in digital or analog form. Spectra are preferably recorded in a storage medium, including, e.g., magnetic, such as floppy disk, tape, or hard disk; optical, such as CD-ROM or laser-disc; or, ROM-CHIPS.

The mass spectrum of a given sample typically provides information on protein intensity, mass to charge ratio, and molecular weight. In preferred embodiments of the invention, the molecular weights of proteins in the sample are used as a matching criterion to query a database. The molecular

weights are calculated conventionally, e.g., by subtracting the mass of the ionizing proton for singly-charged protonated molecular ions, by multiplying the measured mass/charge ratio by the number of charges for multiply-charged ions and subtracting the number of ionizing protons.

Various databases are useful in accordance with the present invention. Useful databases include, databases containing genomic sequences, expressed gene sequences, and/or expressed protein sequences. Preferred databases contain nucleotide sequence-derived molecular masses of proteins present in a known organism, organ, tissue, or cell-type. There are a number of algorithms to identify open reading frames (ORF) and convert nucleotide sequences into protein sequence and molecular weight information. Several publicly accessible databases are available, including, the SwissPROT/TrEMBL database (http://www.expasy.ch).

Typically, a mass spectrometer is equipped with commercial software that identifies peaks above a certain threshold level, calculates mass, charge, and intensity of detected ions. Correlating molecular weight with a given output peak can be accomplished directly from the spectral data, i.e., where the charge on an ion is one and the molecular weight is therefore equal to the numerator value minus the mass of the ionizing proton. However, protein ions can be complexed with various counter-ions and adducts, such as N, C, and K'. In such a case, it would be expected that a given protein ion would exhibit multiple peaks, such as a triplet, representing different ionic states (or species) of the same protein. Thus, it may be necessary to analyze and process spectral data to determine families of peaks arising from the same protein. This analysis can be carried out conventionally, e.g., as described by Mann et al., anal. Chem., 61:1702-1708 (1989).

In matching a molecular mass calculated from a mass spectrometer to a molecular mass predicted from a database, such as a genomic or expressed gene database, post-translation processing may have to be considered. There are various processing events which modify protein structure, including, proteolytic processing, removal of N-terminal methionine, acetylation, methylation, glycosylation, phosphorylation, etc.

A database can be queried for a range of proteins matching the molecular mass of the unknown. The range window can be determined by the accuracy of the instrument, the method by which the sample was prepared, etc. Based on the number of hits (where a hit is match) in the spectrum, the unknown protein or peptide is identified or classified.

Methods of identifying one or more NPP by mass spectrometry are useful for detection in human plasma. Exemplary techniques are described in U.S. Patent Applications 02/0060290, 02/0137106, 02/0138208, 02/0142343, 02/0155509.

Diagnostic and Prognostic Uses

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics as further described herein.

The invention provides diagnostic and prognostic assays for detecting NPP nucleic acids and proteins, as further described. Also provided are diagnostic and prognostic assays for detecting interactions between NPP and target molecules, particularly natural agonists and antagonists.

The present invention provides methods for identifying polypeptides that are differentially expressed between two or more samples. "Differential expression" refers to differences in the quantity or quality of a polypeptide between samples. Such differences could result at any stage of protein expression from transcription through post-translational modification. For example, using protein array methods, two samples are bound to affinity spots on different sets of adsorbents and recognition maps are compared to identify polypeptides that are differentially retained by the two sets of adsorbents. Differential retention includes quantitative retention as well as qualitative differences in the polypeptide. For example, differences in post-translational modification of a protein can result in differences in recognition maps detectable as differences in binding characteristics (e.g., glycosylated proteins bind differently to lectin adsorbents) or differences in mass (e.g., post-translational cleavage products). In certain embodiments, an adsorbent can have an array of affinity spots selected for a combination of markers diagnostic for a disease or syndrome.

Differences in polypeptide levels between samples (e.g., plasma samples) can be identified by exposing the samples to a variety of conditions for analysis by desorption spectrometry (e.g., mass spectrometry). Unknown proteins can be identified by detecting physicochemical characteristics (e.g., molecular mass), and this information can be used to search databases for proteins having similar profiles.

Preferred methods of detecting a NPP utilize mass spectrometry techniques. Such methods provide information about the size and character of the particular NPP isoform that is present in a sample, e.g., a biological sample submitted for diagnosis or prognosis. Mass spectrometry techniques are detailed in the section titled "Detection of NPPs by mass spectrometry". The invention provides a method of detecting a NPP in a biological sample comprising the steps of: fractionating a biological sample (e.g., serum, lymph, cerebrospinal fluid, cell lysate of a particular tissue) by at least one chromatographic step; subjecting a fraction to mass spectrometry; and optionally comparing the characteristics of peptide species observed in mass spectrometry with known characteristics of NPPs.

One embodiment of the present invention involves a method of use (e.g., a diagnostic or prognostic assay) wherein a molecule of the present invention (e.g., a NPP, NPP nucleic acid, or NPP antibody) is used, for example, to diagnose or prognose a disorder in which any of the aforementioned

NPP activities is indicated. In another embodiment, the present invention involves a method of use wherein a molecule of the present invention is used, for example, for the diagnosis or prognosis of subjects, preferably a human subject, in which any of the aforementioned activities is pathologically perturbed. In a preferred embodiment, the methods of use involve administering to a subject, preferably a human subject, a molecule of the present invention for the diagnosis or prognosis. In another embodiment, the methods of use involve administering to a human subject a molecule of the present invention.

For example, the invention encompasses a method of determining whether NPP is expressed within a biological sample comprising: a) contacting said biological sample with: i) a polynucleotide that hybridizes under stringent conditions to a NPP nucleic acid; or ii) a detectable polypeptide (e.g. antibody) that selectively binds to a NPP; and b) detecting the presence or absence of hybridization between said polynucleotide and an RNA species within said sample, or the presence or absence of binding of said detectable polypeptide to a polypeptide within said sample. Detection of said hybridization or of said binding indicates that said NPP is expressed within said sample. Preferably, the polynucleotide is a primer, and wherein said hybridization is detected by detecting the presence of an amplification product comprising said primer sequence, or the detectable polypeptide is an antibody.

In certain embodiments, detection involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegren et al. (1988) Science 241:1077-1080; and Nakazawa et al. (1994) PNAS 91:360-364), the latter of which can be particularly useful for detecting point mutations in the NPP-encoding-gene (see Abravaya et al. (1995) Nucleic Acids Res. 23:675-682).

Also envisioned is a method of determining whether a mammal, preferably human, has an elevated or reduced level of expression of a NPP, comprising: a) providing a biological sample from said mammal; and b) comparing the amount of a NPP or of a NPP-encoding RNA species within said biological sample with a level detected in or expected from a control sample. An increased amount of said NPP or said RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has an elevated level of NPP expression, and a decreased amount of said NPP or said RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has a reduced level of expression of a NPP.

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to

diagnostic assays for determining NPP and/or nucleic acid expression as well as NPP activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NPP expression or activity. The invention also provides for prognostic assays for determining whether an individual is at risk of developing a disorder associated with a NPP, nucleic acid expression or activity. For example, mutations in a NPP-encoding gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NPP polypeptide expression or activity.

The term "biological sample" is intended to include tissues, cells and biological fluids isolated from an individual, as well as tissues, cells and fluids present within an individual. That is, the detection methods of the invention can be used to detect a NPP mRNA, protein, or genomic DNA in a biological sample in vitro as well as in vivo. Preferred biological samples are biological fluids such as lymph, cerebrospinal fluid, blood, and especially blood plasma. For example, in vitro techniques for detection of a NPP-encoding mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detection of a NPP polypeptide include mass spectrometry, enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. In vitro techniques for detection of a NPP-encoding genomic DNA include Southern hybridizations. Furthermore, in vivo techniques for detection of a NPP polypeptide include introducing into an individual a labeled anti-NPP antibody.

In preferred embodiments, the subject methods can be characterized by generally comprising detecting, in a tissue sample of the individual (e.g. a human patient), the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding one of the subject NPP polypeptide or (ii) the mis-expression of a NPP gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from the NPP gene, (ii) an addition of one or more nucleotides to the gene, (iii) a substitution of one or more nucleotides of the gene, (iv) a gross chromosomal rearrangement or amplification of the gene, (v) a gross alteration in the level of a messenger RNA transcript of the gene, (vi) aberrant modification of the gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene, and (viii) reduced level of expression, indicating lesion in regulatory element or reduced stability of a NPP-related transcript.

In yet another exemplary embodiment, aberrant methylation patterns of a NPP-encoding nucleic acid can be detected by digesting genomic DNA from a patient sample with one or more restriction endonucleases that are sensitive to methylation and for which recognition sites exist in the NPP gene (including in the flanking and intronic sequences). See, for example, Buiting et al. (1994)

Human Mol Genet 3:893-895. Digested DNA is separated by gel electrophoresis, and hybridized with probes derived from, for example, genomic or cDNA sequences. The methylation status of the NPP gene can be determined by comparison of the restriction pattern generated from the sample DNA with that for a standard of known methylation.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with an agent capable of detecting a NPP, mRNA, or genomic DNA, such that the presence of said NPP, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of a NPP, mRNA or genomic DNA in the control sample with the presence of a NPP, mRNA or genomic DNA in the test sample. The invention also encompasses kits for detecting the presence of a NPP, mRNA or genomic DNA in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting a NPP, mRNA or genomic DNA in a biological sample; means for determining the amount of a NPP in the sample; and means for comparing the amount of NPP in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NPP or nucleic acid.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either a NPP or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a NPP, or interaction of a NPP with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants and by any immobilization protocol described herein. Alternatively, the complexes can be dissociated from the matrix, and the level of NPP binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a NPP or target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NPP or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NPP or target molecules but which do not interfere with binding of the NPP to its target molecule can be derivatized to the wells of the plate, and unbound target or NPP trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NPP or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the NPP or target molecule.

Pharmaceutical Compositions

When polypeptides of the present invention are expressed in soluble form, for example as a secreted product of transformed yeast or mammalian cells, they can be purified according to standard procedures of the art, including steps of ammonium sulfate precipitation, ion exchange chromatography, gel filtration, electrophoresis, affinity chromatography, according to, e.g., "Enzyme Purification and Related Techniques," Methods in Enzymology, 22:233-577 (1977), and Scopes, R., Protein Purification: Principles and Practice (Springer-Verlag, New York, 1982) provide guidance in such purifications. Likewise, when polypeptides of the invention are expressed in insoluble form, for example as aggregates or inclusion bodies, they can be purified by appropriate techniques, including separating the inclusion bodies from disrupted host cells by centrifugation, solublizing the inclusion bodies with chaotropic and reducing agents, diluting the solubilized mixture, and lowering the concentration of chaotropic agent and reducing agent so that the polypeptide takes on a biologically active conformation. The latter procedures are disclosed in the following references: Winkler et al, Biochemistry, 25: 4041-4045 (1986); Winkler et al, Biotechnology, 3: 992-998 (1985); Koths et al, U.S. patent 4,569,790; and European patent applications 86306917.5 and 86306353.3.

Compounds capable of detecting a NPP or NPP biological activity, including small molecules, peptides, NPP nucleic acid molecules, and anti-NPP antibodies of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutical substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose.

pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL® (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, or thimerosal. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Where the active compound is a protein, e.g., a NPP, sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and other required ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral administration, the compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer. Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated

are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the compounds are formulated into ointments, salves, gels, or creams as generally known in the art. Most preferably, compound is delivered to a subject by intravenous injection.

In one embodiment, the compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

Cardiovascular disorder therapy

A number of agents are useful for the treatment and prevention of cardiovascular disorders. Such agents may be used advantageously in combination with a NPP-related diagnosis or prognosis.

For example, cell cycle inhibitors and proto-oncogenes (Simari and Nabel, Semin. Intervent. Cardiol. 1:77-83 (1996)); NO (nitric oxide) donor drugs; pro-apoptotic agents such as bcl-x (Pollman et al., Nature Med. 2:222-227 (1998)); herpes virus thymidine kinase (tk) gene and systemic ganciclovir (Ohno et al., Science 265:781-784 (1994); Guzman et al., Proc. Natl. Acad. Sci. USA 91:10732-10736 (1994); Chang et al., Mol. Med. 1:172-181 (1995); and Simari et al., Circulation 92:1-501 (1995)) have been exploited to treat atherosclerosis, restinosis and neointimal smooth muscle proliferation.

Anti-thrombotic agents useful in combination with the compositions of the invention include, for example, inhibitors of the IIb/IIIa integrin; tissue factor inhibitors; and anti-thrombin agents. An antiarrhythmic agent, such as a local anesthetic (class I agent), sympathetic antagonist (class II agent), antifibrillatory agent (class III agent) calcium channel agent (class IV agent) or anion antagonist (class V agent) as described in Vukmir, Am. J. Emer. Med. 13:459-470 (1995); Grant, PACE 20:432-444 (1997); Assmann I., Curr. Med. Res. Opin. 13:325-343 (1995); and Lipka et al., Am. Heart J. 130:632-640 (1995) may also be used. Examples of class I agents include: procainamide; quinidine or disopyramide; lidocaine; phenytoin; tocainide or mexiletine; encainide; flecainide; lorcainide; propafenone (III) or moricizine. Sympathetic antagonists include: propranolol, esmolol, metoprolol,

atenelal, or acebutolol. Examples of antifibrillatory agents are bretylium, amiodarone, sotalol (II) or N-acetylprocainamide. Class IV agents include verapamil, diltiazem, and bepridil, and anion antagonists such as alinidine.

Congestive heart failure therapeutic agents include TNF inhibitors such as Embrel.TM. (Immunex Corp.; Seattle, Wash.), TBC11251, or an ACE (angiotensin converting enzyme) inhibitor, such as Natrecor (nesiritide; Scios, Inc.). Angiogenic agents, for example, recombinant VEGF isoforms, such as rhVEGF developed by Genentech; a nucleic acid molecule encoding the 121 amino acid isoform of VEGF (BioByPass.TM.; GenVec/Parke Davis); or a nucleic acid encoding VEGF-2 (Vascular Genetics, Inc.); FIBLAST.TM., a recombinant form of FGF-2 being developed by Scios, Inc. (Mountain View, Calif.) and Wyeth Ayerst Laboratories (Radnor, Pa.), GENERX.TM., or an adenoviral gene therapy vector encoding FGF-4 developed by Collateral Therapeutics (San Diego, Calif.) and Schering AG (see Miller and Abrams, Gen. Engin. News 18:1 (1998)), are also useful in combination with the NPP-related compositions of the invention. Finally, calcium antagonists, such as amlodipine (Marche et al., Int. J. Cardiol. 62(Suppl.):S17-S22 (1997); Schachter, Int. J. Cardiol. 62(Suppl.):S85-S90 (1997)); nicardipine; nifedipine; propanolol; isosorbide dinitrate; diltiazem; and isradipine (Nayler (Ed.) Calcium Antagonists pages 157-260 London: Academic Press (1988); Schachter, Int. J. Cardiol. 62(Suppl.):S9-S15 (1997)) are also advantageous therapeutic agents for cardiovascular disorders.

References cited in the specification are incorporated herein in their entireties. Having generally described this invention, a further understanding can be obtained by reference to certain specific examples which are provided herein for purposes of illustration only, and are not intended to be limiting unless otherwise specified.

EXAMPLES

Example 1: Collection of plasma samples from experimental and control populations

Subjects enrolled in the Duke Databank for Cardiovascular Disease were selected on the basis of coronary artery disease (CAD). A total of 241 CAD patients and control individuals were further matched for gender, age, and ethnicity and individuals with plasma abnormalities were excluded. A set of 53 CAD patients and a set of 53 control individuals were established. Six liters of plasma were pooled from each set. An aliquot of plasma was retained from each individual, thus allowing a positive result in the pooled sample to be confirmed for each member of the population. Such confirmation is valuable to erase possible confounding effects of an individual with an aberrant level of a specific polpeptide that is not related to a cardiovascular disorder.

Example 2: Characterization of low molecular weight NPP levels in experimental and control plasma

Two and a half liters of pooled plasma from each population were subjected to separation by multiple chromatography steps according to the Microprot.TM process as follows:

Step 1: HSA/IgG depletion

125 ml frozen plasma were defrost and filtered on 0.45 µm sterile filter in a sterile hood.

Filtrate was injected on two inline columns of respectively 300 ml of HSA ligand Sepharose fast Flow column (Amersham, Upsala, Sweden), 5cm ID, 15 cm length; and 100 ml Protein G Sepharose fast Flow column (Amersham, Upsala, Sweden), 5 cm ID, 5 cm length.

Columns were equilibrated and washed with 50 mM PO4 buffer, pH 7.1, 0.15M NaCl. Flow rate was 5 ml/min.

Non-retained fraction (350 ml) was frozen until second step. Twenty runs were performed.

Step 2: Gel Filtration / Reverse Phase Capture step

Sample from step 1 was defrosted and filtered on 0.45 µm sterile filter in a sterile hood.

Filtrate was injected on two in line gel filtration columns: 2 X 9.5 litres Superdex 75 (Amersham, UK) column, 14 cm ID, 62 cm length. Column was equilibrated with 50mM PO4 buffer pH 7.4, 0.1 M NaCl, 8M urea. Hydrophobic impurities were retained on a reverse phase precolumn: 150 ml PLRPS (Polymer Labs, UK). Precolumn was switched for sample injection. Gel filtration was performed at a flow rate of 40 ml/min.

Low molecular weight proteins (<20 kDa) were oriented to in line reverse phase capture column: 50 ml PLRPS 100 angstroms (Polymer labs, UK). The three-way valve controlling injection on PLRPS column was switched at a cut-off of 33 mAU (280 nm) to send gel filtration eluate into reverse phase capture column. This cut-off value was established by first using SDS-PAGE to provide an estimated range of OD values and by subsequently evaluating three cut-off values (high, median and low values of OD range). The final cut-off value was chosen to maximize the low molecular weight protein obtained, with a low molecular protein proportion of at least 85%. Low molecular weight proteins and peptides were eluted from reverse phase capture PLRPS column by one column volume gradient of 0.1% TFA, 80% CH3CN in water.

Eluate fractions (50 ml) were frozen until next step. Twenty runs were performed. At the end of this step, all reverse phase eluates were defrosted, pooled (1 liter) and shared in 7 polypropylene containers (143 ml). Containers were kept at –20°C until use for next step.

Step 3: Cation Exchange

Sample from step 2 (147 ml) was defrosted and mixed with an equal volume of cation exchange buffer A (Gly/HCl buffer 50 mM, pH 2.7, urea 8M).

Sample was injected on a 100 ml Source 15S column (Amersham, Upsala, Sweden), 35 mm ID, 100 mm length. Column was equilibrated and washed with buffer A. Flow rate was 10 ml/min.

Proteins and peptides were eluted with step gradient from 100% buffer A until 100 % buffer B (buffer A containing 1M NaCl):

3 column volumes 7.5% B (75 mM NaCl)

3 column volumes 10% B (100 mM NaCl)

3 column volumes 17.5% B (175 mM NaCl)

2 column volumes 22.5% B (225 mM NaCl)

2 column volumes 27.5% B (275 mM NaCl)

2 column volumes 100% B (1 M NaCl)

45 to 60 fractions were collected based on peak. Seven runs were conducted. After 7 runs were achieved, fractions were pooled intra and inter run in order to obtain 18 fractions. Fractions were kept at –20°C until use for next step.

Step 4: Reduction/Alkylation and Reverse Phase HPLC Fractionation 1

After adjusting the pH to 8.5 with concentrated Tris-HCl, each of the 18 cation exchange fractions was reduced with dithioerythritol (DTE, 30 mM, 3 hours at 37°C) and alkylated with iodoacetamid (120 mM, 1 hour 25°C in the dark). The latter reaction was stopped with the addition of DTE (30 mM) followed by acidification (TFA, 0.1 %). The fractions were then injected on an Uptispher C8, 5 um, 300 angstroms column (Interchim, France), 21 mm ID, 150 mm length. Injection was performed with a 10 ml/min flow rate.

C8 column was equilibrated and washed with 0.1 % TFA in water (solution A). Proteins and peptides were eluted with a biphasic gradient from 100% A until 100% B (0.1% TFA, 80% CH3CN in water) in 60 min. Flow rate was 20 ml/min. Thirty fractions of 40 ml were collected.

Based on the measured optical density (OD) at 280 nm of each fraction, which reflects the protein concentration in that fraction, aliquots of similar protein content were created for each fraction.

All aliquots were frozen and kept for further use except one per fraction which was dried with a Speed Vac (Savant, Fischer, Geneva) after addition of 500 ul 10% glycerol in water in each fraction, in order to prevent excess drying. Dried fractions were kept at -20°C until use for next step.

Step 5: Reverse Phase HPLC Fractionation 2

Dried samples from step 4 were resuspended in 1 ml of solution A (0.03% TFA in water) and injected on a Vydac LCMS C4 column, 5 micrometers, 300 angstroms (Vydac, USA), 4.6 mm ID, 150 mm length. Flow rate was 0.8 ml/min.

C4 column was equilibrated and washed with solution A and proteins and peptides were eluted with a biphasic gradient adapted to elution position of the sample in Reverse Phase HPLC

Fractionation 1. Intact mass data were acquired using Electrospray Ion Trap Mass spectrometry. Sixteen different gradients were used with a CH3CN concentration range minus and plus 5% CH3CN of RP1 fraction corresponding solvent concentration. For proteins eluted in RP1 with a solvent concentration equal to or greater than 30 % CH3CN, the starting elution conditions for the RP2 gradient was set, in CH3CN percentage, at the RP1 elution concentration minus 30%. Twenty-four eluted fractions were collected in a deep well plate, adopting optimized different collection configurations designed for optimal SpeedVac concentration and further robotic treatment.

Step 6: Mass detection

About 13,000 fractions were collected following reverse phase HPLC fractionation 2 into 96-well deep well plates (DWP). A small proportion (2.5%) of the volume was diverted to online analysis using LC-ESI-MS (Bruker Esquire). Aliquots of undigested proteins were mixed with MALDI matrices, and spotted on MALDI plates together with mass calibration standards and sensitivity standards. Automated spotting devices (Bruker MALDI sample prep. robots) were used. Two different MALDI matrices were employed: sinapic acid (SA), also known as sinapinic acid, trans-3,5-dimethoxy-4-hydroxycinnamic acid, and alpha-cyano-4-hydroxycinnamic acid (HCCA). MALDI plates were subjected to mass detection using Bruker Reflex III MALDI MS apparati. The 96-well plates were stored at +4 C.

96-well plates (DWP) were recovered and subjected to two sequential concentration steps. Volumes were concentrated from 0.8 ml to about 50 microl per well by drying with a SpeedVac, and then resolubilized to *ca.* 200 microl and reconcentrated to about 50 microl per well, and stored at +4 C. Proteins were then digested by re-buffering, adding trypsin to the wells, sealing and incubating the plates at 37 C for 12 hours, followed by quenching (addition of formic acid to bring the pH down to 2.0). The concentration of trypsin to be added to the wells was adjusted based on the OD at 280 nm recorded for each particular fraction. This ensured an optimal use of trypsin and a complete digestion of the most concentrated fractions. Automated spotting devices (Bruker MALDI sample prep. robots) were used to deposit a volume from each well, pre-mixed with a HCCA matrix onto a MALDI plate together with sensitivity and mass calibration standards. MALDI plates were analyzed using a Bruker Reflex III MALDI MS device. Contents from each well of the 96 well plates were analyzed with LC-ESI-MS-MS Bruker Esquire ESI Ion-Trap MS devices as described in Example 4.

TABLE 2

TABLE 2 Peptide SEQ ID NO	Peptide Sequence	CEX	Salt	RP1	% B	Run Number
1	SICPSALIKISLER	12	175	16	49.02	95681_10
2	IASSGKTGIQTK	2	75	17	50.95	130115_21
		2	75	17	50.95	130115_21
		2	75	19	54.8	84230_20
		4	75	20	56.71	88169_12
		4	75	20	56.71	88169_12
		4	75	21	58.64	88193_13
		4	75	21	58.64	88193_13
		4	75	21	58.64	88193_13
3	HPNMLTECLLCGK	11	175	8	33.65	88265_14
		11	175	8	33.65	88265_14
		11	175	9	35.7	110344_13
		11	175	10	37.5	88205_09
5	KMPLFIYICTK	17	275	25	68.25	105084_04
		17	275	27	70.17	105128_04
6	SAAHLILLR	3	75	24	64.4	89830_03
		5	75	20	56.71	118038_05
		5	75	23	62.48	118054_06
		6	100	19	54.8	130950_12
		7	100	12	41.34	130868_21
		9	175	12	41.34	105012_13
7	LPTTMLIGR	1	75	18	52.87	88041_09
8	QMVLMSCVLK	14	175	15	47.1	117990_17
		14	175	15	47.1	117990_17
9	QILENQVR	14	175	13	43.27	92590_16
10	CVSSYPTSAEK	9	175	18	52.87	108131_08
11	LCVLIMK	11	175	20	56.71	89770_14
		11	175	20	56.71	89770_14
12	STNAHLGAKR	10	175	9	35.7	89870_20
		11	175	10	37.5	110368_12
13	FGKTDNINCPK	2	75	23	62.48	84198_17
		2	75	23	62.48	84198_17
14	WSPECSSTSIVLR	6	100	8	33.65	130804_08
15	GGNVCGTVANGKQEK	14	175	6	29.8	117854_13
17	MVLDVSDNEMTFSK	15	225	18	52.87	121207_10
18	VMLMIQETNK	1	75	10	37.5	88009_11
		1	75	10	37.5	88009_15
		1	75	11	39.42	88029_17
		1	75	11	39.42	88029_17
		1	75	12	41.34	88045_15
		2	75	9	35.7	84246_18
		2	75	9	35.7	84246_18
		2	75	9	35.7	84246_21
		2	75	9	35.7	84246_23
19	IVHQVSKLFK	17	275	18	52.87	111989_12

TABLE 2

Peptide SEQ ID NO	Peptide Sequence	CEX	Salt	RP1	% B	Run Number
20	LLNNFPYR	8	100	15	47.1	100961_08
21	RPLSSSHIGSPR	6	100	8	33.65	111981_08
22	KGAPLLGK	10	175	20	56.71	110512_08
23	RMNSAFGGR	9	175	13	43.27	130876_22
24	QGSGHIGK	1	75	12	41.34	117938_10
		1	75	19	54.8	88061_05
		1	75	19	54.8	88061_24
		1	75	21	58.64	118046_08
		1	75	21	58.64	118046_10
		1	75	21	58.64	118046_11
		1	75	22	60.56	118058_06
		2	75	21	58.64	130111_24
		2	75	21	58.64	130111_24
		2	75	21	58.64	84250_05
		2	75	21	58.64	84250_08
		2	75	21	58.64	84250_20
		2	75	22	60.56	121271_12
		2	75	23	62.48	84198_17
		2	75	23	62.48	84198_22
		2	75	23	62.48	84198_22
		2	75	23	62.48	84198_23
		2	75	28	72.1	84270_03
		2	75	28	72.1	84270_07
		2	75	28	72.1	84270_10
		3	75	17	50.95	89818_11
		4	75	18	52.87	88145_11
		4	75	22	60.56	88109_06
		4	75	22	60.56	88109_07
		4	75	24	64.4	110204_05
		5	75	25	68.25	100678_09
		7	100	22	60.56	108047_06
		8	100	24	64.4	100997_05
		10	175	22	60.56	117618_05
		10	175	22	60.56	117618_05
		17	275	30	87.5	130508_09
26	LIFVCEASLHPK	9	175	9	35.7	108111_12
27	SGCTNLRSHQQCIR	14	175	24	64.4	92570_10
28	QGWQGNSIGKK	14	175	10	37.5	117918_09
31	ESIYFIIAAMLVATK	1	75	6	29.8	88053_11
32	IFLLGQITSIPDKL	10	175	22	60.56	117618_03
33	KPLKNGSQFS	11	175	5	27.88	110292_08
34	RVITPLIK	2	75	19	54.8	84230_19
36	RHCLLFVCFCK	10	175	26	69	87981_01
37	CHFCLTCSR	13	175	18	52.87	92144_09
38	IPTTFETNL	1	75	13	43.27	88065_21
		2	75	20	56.71	84234_24

TABLE 2 Peptide SEQ ID NO	Peptide Sequence	CEX	Salt	RP1	% B	Run Number
39	STVLSASLHLR	10	175	6	29.8	110304_06
		10	175	6	29.8	110304_06
40	LEVELTFLWPSPPR	3	75	21	58.64	89874_17
41	IFLTMDQLLQN	11	175	20	56.71	89770_05
42	SASLMEIQSKK	1	75	12	41.34	88045_11
43	MKPLVDYK	10	175	23	62.48	87941_05
44	EDLGSKGPK	11	175	17	50.95	89718_11
46	TFSSALFWK	17	275	8	33.65	100774_08
47	QADGTVFSK	2	75	15	47.1	121151_09
48	SILFAFSLYR	2	75	9	35.7	84246_20
		2	75	10	37.5	84178_19
		2	75	12	41.34	118923_10
		3	75	13	43.27	84318_13
		10	175	21	58.64	88001_14
		10	175	21	58.64	88001_15
		10	175	21	58.64	88001_16
49	SSPLDLVCNSSSTSY	9	175	3	20.8	108135_15
50	VKMLHALVLK	14	175	21	58.64	92096_06
51	ADSGLAQSDGK	11	175	22	60.56	89698_06
52	DHEDAWRMFSAR	6	100	5	27.88	130732_11
		6	100	5	27.88	130732_11
		6	100	5	27.88	100730_13
		6	100	5	27.88	100730_13
		6	100	5	27.88	100730_13
		9	175	4	25.96	108067_13
		9	175	5	27.88	108083_10
		9	175	5	27.88	108083_11
		9	175	6	29.8	108091_07
53	CVIFPLNSYGMLLK	3	75	11	39.42	84286_20
		4	75	12	41.34	110212_16
		4	75	12	41.34	88141_15
54	HLKLAISSLLR	2	75	10	37.5	84178_22
55	DSYLNVKR	10	175	11	39.42	110376_13
		10	175	11	39.42	110376_13
		11	175	11	39.42	110380_14
		11	175	11	39.42	110380_15
56	HSELCLAR	16	225	28	72.1	105140_03
		16	225	28	72.1	105140_03
57	CSKTFINTK	11	175	13	43.27	110412_16
		13	175	22	60.56	92112_03
58	NRQTLLLLMSCR	16	225	14	45.2	105068_08
59	YLSDGWIKGYIK	14	175	15	47.1	117990_17
60	DVSSAIPNSVS	13	175	15	47.1	110440_11
61	VSWHKHLLLLR	15	225	10	35.7	118907_10
62	EAEFESTMQK	12	175	13	43.27	130131_19
63	ILMDDFKK	15	225	18	52.87	121207_10

TABLE 2 Peptide SEQ ID NO	Peptide Sequence	CEX	Salt	RP1	% B	Run Number
64	YSEIKEK	15	225	18	52.87	121207_10
65	SRHQEIGCLAR	16	225	25	68.25	130435_10
66	QVQSYHVLGK	16	225	11	39.42	130588_08
67	NPMKIFEK	14	275	24	64.4	117658_03
68	LPNHLLNHR	6	100	12	41.34	100928_08
69	ETLMAAELNMAGIYNGIKGAR	3	75	2	12.5	84314_12
		12	175	3	20.8	121103_07
		17	275	3	20.8	130692_09
		17B	275	2	12.5	130668_10
70	PLTLWSHR	1	75	10	37.5	88009_10
71	SSLVLYVLR	8	100	11	39.42	121095_12
73	CLCTHNGASKYMK	2	75	19	54.8	84230_19
		9	175	21	58.64	108147_07
		9	175	21	58.64	108147_08
74	LGFLFVSETESR	15	225	5	27.88	118851_11
75	ICNIQQAHIHWR	6	100	14	45.2	130910_12
77	CLYSFVFSR	6	100	9	35.7	130820_14
		8	100	10	37.5	121091_08
78	ENVIPSLTVPK	3	75	11	39.42	84286_12
79	KTILEHIPLR	12	175	11	39.42	111973_14
80	KSCVGLTTFY	13	175	21	58.64	92192_08
81	LSAAVRLSAAVR	16	225	12	41.34	100924_13
82	QQHKSASLLR	9	175	5	27.88	108083_08
83	QDHLNISYK	1	75	15	47.1	117998_08
87	YALKCHNLQILHTK	5	75	25	68.25	100678_02
89	FSDDTHRTGR	15	225	20	56.71	121231_04
		16	225	18	52.87	111985_15
90	TAVVSLPR	13	175	22	60.56	117630_12
91	EQLSLLDR	13	175	22	60.56	117630_12
92	AVLDVFEEGTEASAATAVK	13	175	22	60.56	117630_12
93	ITLLSALVETR	13	175	22	60.56	117630_12
94	ILHMLCHLILIR	3	75	27	70.17	110164_02
95	IHQQLALWTWK	13	175	10	37.5	89706_08
96	PEMVVQACSLSY	8	100	23	62.48	100985_08
97	LMYLVFTKASPK	11	175	9	35.7	110344_11
99	WFLRILGSPMGVLSQWGK	11	175	26	69	89766_01
100	GTELLIHHQWPK	3	75	24	64.4	89830_03
101	ALHLDNSAFR	5	75	18	52.87	118018_06
102	NAKISQAPW	10	175	23	62.48	87941_03
103	CWATESNEIHLEIQT	13	175	23	62.48	92128_12
104	LFLDCMLNK	15	225	18	52.87	121207_10
105	LFIFTCVFHK	15	225	5	27.88	118851_11

TABLE 2 Peptide SEQ ID NO	Peptide Sequence	CEX	Salt	RP1	% B	Run Number
106	HCRTNHVLLLLR	2	75	15	47.1	130103_17
		3	75	24	64.4	110120_06
		4	75	13	43.27	110224_11
		4	75	14	45.2	110248_13
		4	75	15	47.1	110256_14
		4	75	15	47.1	110256_14
		4	75	15	47.1	110256_14
		9	175	12	41.34	130860_21

Example 3: Characterization of high molecular weight NPP levels in experimental and control plasma

Two and a half liters of pooled plasma from each population were subjected to separation by multiple chromatography steps according to the Macroprot.TM process as follows:

Step 1: Differential labelling of control and CAD plasma protein samples

The control and CAD samples are first labeled with separate dyes that fluoresce at different wavelengths, thereby exhibiting different colors when appropriately viewed. After labeling, the samples are mixed, and fractionated according to the following protocol. Proteins that are differentially present in the control and CAD samples are thus detectable on a 2 dimensional (2D) gel, after fractionization, as described in U.S. Patent No. 6,043,025.

Depletion by affinity chromatography:

Each 125 ml aliquot was filtered and applied to a 318 ml HSA affinity column. Serum albumin (HSA) is removed from the sample by injecting it (at a linear flow rate of 15 cm per h) onto a chromatographic affinity column previously equilibrated by passing 5 column volumes (CVs) of equilibration buffer (20 mM sodium phosphate, 50 mM sodium chloride, pH 7.1). The column contains a cross-linked agarose matrix (Sepharose – Amersham Biosciences) to which a ligand specific for HSA has been coupled (HSA affinity column – Amersham Biosciences). An HSA binding capacity of 15 mg HSA per ml matrix is respected. A linear flow rate of 70 cm per h is maintained while the non-retained (NR) fraction elutes from the column, as indicated by UV absorbtion increase. This is collected until the UV absorbtion has returned to zero. The NR volume is subsequently injected without buffer change onto a second affinity column with a specificity for immunoglobulin class G (IgG) proteins (ProteinG IgG affinity column – Amersham Biosciences). An IgG binding capacity of 17 mg IgG per ml matrix is respected. The flow rates for injection and elution for this column are identical to those described for the HSA affinity column. This procedure was repeated twenty times to treat the entire volume of 2.5 L. The NR fractions from these runs comprised in total

approximately 8.5 L (considerable dilution is an effect of some chromatography separations) containing 44 g total protein.

Step 2: Fractionation by affinity chromatography

The 20 fractions were pooled and re-aliquoted into 4 equal volumes. These fractions were fractionated in 4 identical runs on a 625ml benzamidine sepharose column (Amersham) previously equilibrated in equilibration buffer (20 mM sodium phosphate, 200 mM NaCl, pH 7.4, 2 CVs). A loading capacity of 7 mg protein per ml matrix is respected and all fractionation is performed using a linear flow rate of 70 cm per h. Following these runs the total volumes and protein contents obtained following pooling for the four fractions were 10.7 L, 9.6 g (B1), 13.8 L, 12.6g (B2) and 7.8 L, 12.7 g (B3).

Further separation of fractions B2 and B3 proceeded on a second affinity column. Urea was added in powder form at 2 mol/l to the fraction B2. Following solubilization of the urea, the fraction was rendered compatible with the second affinity step by buffer change on a G25 matrix (Amersham Biosciences) gel filtration column. The volume of the G25 column was 4 times larger than the sample volume. The protein was eluted in a buffer of 50 mM sodium phosphate, 2 M urea, 50 mM NaCl (pH 7.1). The B2 fraction was then injected onto a 815ml column of Red Sepharose matrix (Amersham Biosciences), previously equilibrated with a buffer of 50 mM sodium phosphate, 50 mM NaCl, 2 M urea (pH 7.1) at a loading of 6 mg protein/ ml matrix. The loading and elution flow rate was 50 cm/ h. Two fractions were obtained from this separation, the NR fraction (B2R1) and a fraction eluted with a buffer of 50 mM sodium phosphate, 2 M NaCl, 2 M urea (pH 7.1) (B2R2). Urea was added at 1 mol/ l to B3 and, following solubilization, the pH was adjusted with concentrated NaOH to pH 7.1. This fraction was then injected onto a column of Red Sepharose equilibrated with a buffer of 50 mM sodium phosphate, 2 M urea (pH 7.1). Loading capacity and flow rates were the same as those used for the separation of the B2 fraction. Two fractions were obtained from this separation, the NR fraction (B3R1) and a fraction eluted with a buffer of 50 mM sodium phosphate, 2 M NaCl, 2 M urea (pH 7.1) (B3R2). Following the two red sepharose runs total volumes and protein contents obtained from the two pooled fractions were 18.9 L, 9.2 g (B2R1), 1.4 L, 1.4 g (B2R2), 8.4 L, 3.2 g (B3R1) and 4.5 L, 6.8 g (B3R2).

Step 3: Strong ion exchange chromatography:

The fractions B1, B2R1, B3R1 and B3R2 were prepared for strong cation exchange (SCX) chromatrography by buffer exchange on a column of G25 matrix. The fraction B2R2 was prepared for strong anion exchange similarly. To respect the matrix loading capacity fractionation by SCX on a column of 1113 ml was performed in 5 (B1), 5 (B2R1), 2 (B3R1) and 4 (B3R2) identical runs. Fractionation of B2R2 on SAX on a column of 314 ml was performed in 2 identical runs. Separation

of these samples produced 6 fractions for B1 (named B1C1 etc.), 4 fractions for B2R1 (B2R1C1 etc.), 5 fractions for B2R2 (B2R2A1 etc.), 5 fractions for B3R1 (B3R1C1 etc.) and 5 fractions for B3R2 (B3R2C1 etc.)- 25 fractions in total to be carried forward to separation by electrophoresis. Volumes (L) and protein quantities (mg) for these samples, following pooling of the fractions from different runs were as follows: B1C1(31.1; 698), B1C2(5.5; 484), B1C3(5.1; 507), B1C4(4.9; 2057), B1C5(5.06; 800), B1C6(9.06; 1342), B2R1C1(26.6; 1045), B2R1C2(5.5; 895), B2R1C3(6.6; 444), B2R1C4(5.3; 293), B2R2A1(3.1; 97), B2R2A2(0.7; 117), B2R2A3(0.9; 242), B2R2A4(0.4; 90), B2R2A5(0.6; 56), B3R1C1(10.5; 248), B3R1C2(2.2; 208), B3R1C3(1.9; 270), B3R1C4(2.1; 436), B3R1C5(0.7; 232), B3R2C1(12.0; 397), B3R2C2(4.4; 178), B3R2C3(3.8; 256), B3R2C4(5.6; 1241) and B3R2C5(5.2; 478).

Step 4: Preparation for further electrophoretic separation

Following SCX or SAX fractionation the 25 samples were subjected to buffer change on appropriate sized G25 columns using SCX equilibration buffer. They were then concentrated also by SCX, this time being eluted in one peak by an elution step of 1 M NaCl, 50 mM glycine, pH 9.5. Each fraction was then applied to a commercially available isoelectric focusing (IEF) apparatus, the Rotofor® (Bio-Rad).

Step 5: Separation by two-dimentional (2D) gel electrophoresis

Protein samples were solubilized in 9M urea, 2% NP-40, 2% of a pH 8-10.5 ampholyte mixture and 1% dithiothreitol (DTT) and analyzed using 2-D electrophoresis. Corresponding fractions from both control and CAD samples were each applied to the same gel and separated on the basis of isoelectric point and electrophoretic mobility. Thus, a plurality of gels was generated, each representing a separate fraction obtained by the fractionation process described above.

The gels were then scanned to obtain images corresponding to the first and the second fluorophores (i.e., indicating control and CAD samples) using two different wavelengths. Subsequently, the gels were further stained with the SyproRuby dye, and imaged for this dye. Image analysis was conducted using the Progenesis software (Nonlinear Dynamics, Durham, NC, USA) to analyze for each gel the images derived from the two differential fluorophores and from the SyproRuby stain. The images obtained using the first and the second fluorophores were then analysed to quantify any differential expression between the control sample and the CAD sample. The comparison of the images obtained using the first fluorophore and the SyproRuby stain allowed a relationship to be established for each spot, the first fluorophore providing the quantitation information and the SyproRuby, being more sensitive, providing direction for the gel-cutter robot. Similarly, a comparison was conducted between the images obtained using the second fluorophore and the SyproRuby stain.

Gel spots corresponding to proteins were excised from the 2D gel as described in U.S. Patent 6,278,794. A spot pick list was generated based on protein concentration and differential intensity. A spot-picking device was used to isolate the selected protein from gel by removing a portion of the polyacrylamide gel containing the selected protein. The gel plugs containing the proteins of interest were washed and buffered. A detector measuring the optical density (OD) at 280 nm is operably connected to a computer that calculates protein concentration for each fraction and calculates the amount of trypsin to be added to the fraction. Based on said calculation, the computer directs a robotic device to dispense the corresponding amount of trypsin to the fraction for in-gel digestion.

Following in-gel digestion of proteins, proteins were prepared for MS-MALDI or liquid chromatography, electrospray ionization (LC-ESI)-MS-MS. Proteins were extracted from the gel and prepared for MALDI using HCCA (trans-3,5-dimethoxy-4-hydroxycinnamic acid) as MALDI matrix or for LC-ESI-MS-MS by pooling proteins. MALDI-MS was carried out, generating one peptide mass fingerprint (PMF) per isolated spot, and LC-ESI-MS-MS was carried out, generating a fragmentation spectrum.

Two different MALDI matrices were employed: sinapic acid (SA), also known as trans-3,5-dimethoxy-4-hydroxycinnamic acid, and alpha-cyano-4-hydroxycinnamic acid (HCCA). MALDI plates were subjected to mass detection using Bruker Reflex III MALDI MS apparati.

Automated spotting devices (Bruker MALDI sample prep. robots) were used to deposit a volume from each well, pre-mixed with a HCCA matrix onto a MALDI plate together with sensitivity and mass calibration standards. MALDI plates were analyzed using a Bruker Reflex III MALDI MS device. Contents from each well of the 96 well plates were analyzed with LC-ESI-MS-MS Bruker Esquire ESI Ion-Trap MS devices as described in Example 4.

TABLE 3

TABLE 3 Peptide SEQ ID NO	Peptide Sequence	Benzamidine Red Sepharose	SCX SAX	Rotofor	Run Number
4	RLGHGIDAQ	B2R1	C3	R3	183455_02
	RLGHGIDAQ	B2R1	C3	R3	183455_06
16	HTNYFLKNHS	B3R1	C4	R3	151599_16
25	KAVNALAHK	B2R1	C3	R10	183527_14
29	LVPVLQI	B3R2	C5	R9	165081_17
30	TEGLTLLQLV	B3R2	C5	R9	165081_17
35	LGTVSLTH	B3R2	C5	R8	183859_22
45	CLLLRGHYSAMR	B1	C2	R6	183819_09
72	MPGILYNK	B1	C6	R12	178908_21
76	EQNKILSNLEIER	B3R1	C4	R5	151507_13
84	INEKIFCGHK	B3R1	C5	R10	153528_11
85	CTSVDHTPIR	B3R1	C4	R6	152197_01
86	SHLNVQSEKVK	B3R1	C2	R7	152145_08

TABLE 3 Peptide SEQ ID NO	Peptide Sequence	Benzamidine Red Sepharose	SCX SAX	Rotofor	Run Number
88	FCKFSLLISSSTR	B1	C4	R17	183711_09
90	TAVVSLPR	B2R1	C4	R6	180092_11
	TAVVSLPR	B3R1	C3	R18	162590_09
91	EQLSLLDR	B2R1	C4	R6	180092_11
93	ITLLSALVETR	B2R1	C4	R6	180092_11
98	EDNTAEYEPCALR	B1	C4	R17	179852_01

Example 4: Detection and identification of NPPs

Separated fractions and excised gel spots are subjected to mass spectrometry (both matrix-assisted laser desorption/ionization (MALDI) and MS-MS) for separation and detection.

Intact mass data, Peptide Mass Fingerprints and peptide sequence data were integrated for protein identification and characterization. Proteins were identified using Mascot software (Matrix Science Ltd., London, UK), and results from peptide identification were checked by manual analysis of the spectra.

NPP tryptic peptides separated according to the method of Example 2 include: SEQ ID NOs:1-3, 5-15, 17-24, 26-28, 31-34, 36-44, 46-71, 73-75, 77-83, 87, 89-97, and 99-106. NPP tryptic peptides separated according to the method of Example 3 include: SEQ ID NOs:4, 16, 25, 29, 30, 35, 45, 72, 76, 84, 85, 86, 88, 90, 91, 93, and 98.